State-of-the-Art Report

Naturalistic Decision Making: Implications for Design

Gary Klein, Ph.D.
Klein Associates Inc.

April 1993

20081009161

13. ABSTRACT (Maximum 200 words)

Recent years have witnessed strong progress in understanding how people make decisions in operational settings. The emerging field of Naturalistic Decision Making (NDM) is at a point to afford system developers (including design engineers, human factors engineers, ergonomics specialists) different tools and methods for designing interfaces/systems that will better support decision making in those settings. Decision requirements can be identified from the early conceptual design phase through redesign.

The NDM framework attempts to describe the way in which people handle difficult conditions within the context of the overall setting or task. This SOAR describes various decision strategies used by individuals and teams to assess a situation, diagnose a problem, and select a course of action. The impact of stress upon these strategies is also considered. To help understand what people are thinking as they perform difficult tasks, the procedures for conducting Cognitive Task Analyses to examine design requirements are also examined.

State-of-the-Art Report

NATURALISTIC DECISION MAKING:
IMPLICATIONS FOR DESIGN

Gary Klein, Ph.D.

Klein Associates Inc.
Fairborn, OH

April 1993

For

Helen Gail

Devorah

Rebecca

ABOUT CSERIAC

The Crew System Ergonomics Information Analysis Center (CSERIAC) is the gateway to worldwide sources of up-to-date human factors information and technologies for designers, engineers, researchers, and human factors specialists. CSERIAC provides a variety of products and services to government, industry, and academia promoting the use of ergonomics in the design of human-operated equipment and manned systems.

CSERIAC's primary objective is to acquire, analyze, and disseminate timely information on ergonomics. On a cost-recovery basis, CSERIAC will:

- Distribute ergonomic technologies and publications
- Perform customized bibliographic searches and reviews
- Prepare state-of-the-art reports and critical reviews
- Conduct specialized analyses and evaluations
- Organize and/or conduct workshops and conferences

CSERIAC is a Department of Defense organization sponsored by the Defense Technical Information Center. It is managed by the Human Engineering Division of the Armstrong Laboratory and is operated by the University of Dayton Research Institute.

To obtain further information, contact:

CSERIAC Program Office
AL/CFH/CSERIAC, Building 248
2255 H Street
Wright-Patterson AFB, OH 45433-7022
(513) 255-4842
DSN 785-4842
FAX (513) 255-4823
DSN FAX (513) 785-4823

EXECUTIVE SUMMARY

PURPOSE: To describe the current state of knowledge about the way people make decisions in operational settings. This SOAR is written for system developers, including design engineers, human factors engineers, ergonomics specialists, and others who try to design systems, subsystems, and interfaces that will support better decision making.

The problem addressed by this SOAR is that system developers usually aren't given details about how the people operating the system will use it to make difficult judgments and decisions. The SOAR explains how to obtain the decision requirements, and how to incorporate them into the design process. The SOAR also describes tools for identifying and using decision requirements. The SOAR is intended to show developers how to use decision requirements tools and methods to clarify system features and to design system interfaces that are easier to use at critical times.

BOUNDS: This SOAR does NOT provide a detailed review of the decision-making literature of the past 35 years, since that research literature has focused on laboratory studies that usually control out some of the important variables found in operational settings, e.g., high stakes, changing conditions, time pressure, and highly experienced participants. The SOAR does emphasize the recent work in the Naturalistic Decision Making (NDM) framework, which attempts to understand and describe the way people handle difficult conditions within the context of the overall tasks and conditions.

DECISION STRATEGIES: This SOAR describes the strategies people use for situation assessment and for diagnosing a problem, as well as strategies people use to select a course of action. The SOAR explains how stress affects the decision making of both individuals and teams.

COGNITIVE TASK ANALYSIS: This SOAR explains how cognitive probes can be used to understand what people are thinking about as they perform difficult tasks. Four general procedures for Cognitive Task Analysis, contrasting the strengths and weaknesses of each, and showing how a Cognitive Task Analysis would be used to define the decision requirements, will be described.

COGNITIVE SYSTEMS ENGINEERING: There are many ways to take cognitive processes into account during design, e.g., reducing

working memory load as well as making it easier to direct attention. The use of decision requirements is one more way to build systems and interfaces that reflect cognitive processes. Decision requirements can be identified and used throughout the decision process, from early conceptual design to system specification, test and evaluation, and redesign.

CONCLUSIONS: This report describes how people make decisions in operational settings. System designers can identify decision requirements and use these requirements to support the difficult portions of a mission. Design engineers are frequently asked to work on systems, subsystems, and interfaces without being given the information about how the people operating the system will use it to make decisions. The designers may be told the task, e.g., protect an aircraft from threats. But they usually aren't told the specific decision, e.g., for self protection, the operator will be timing out the nearest threats in balance to the nearest friendly interceptor. And designers are rarely given information about the nature of the decision strategy—how the operator will likely use certain rules of thumb and comparisons.

The field of Naturalistic Decision Making tries to develop tools for anticipating how operators will use a system to make difficult decisions. If design engineers are given these tools for anticipating how operators will use a system to make difficult decisions, the result should be more robust interfaces and better decision support. Therefore, the primary value of NDM is to define the decision requirements for a system being developed. These decision requirements can clarify information needs and enable designers to generate interface formats, and to make tradeoffs.

NDM describes the strategies people use for situation assessment, and for diagnosing a problem, as well as strategies that people use to select a course of action. The report examines how stress affects decision making of individual operators and also team decision making. The report also explains how a designer would go about supporting naturalistic decision strategies.

In assessing the decision requirements for a system, design engineers need to understand how the operators think about the task. Cognitive Task Analysis can capture the operators' thought processes. Four methods of Cognitive Task Analysis are presented, along with guidelines for adopting and applying each.

In the past few years, there has been strong progress in understanding the way people make decisions in operational contexts. The field is now at a point where system developers can represent decision requirements during early conceptual design, preparation of specifications, Test and Evaluation, and redesign. The Naturalistic Decision Making framework can provide tools for ensuring that systems will satisfy the decision requirements of operational settings.

ACKNOWLEDGMENTS

ABOUT THE AUTHOR

Dr. Gary Klein is on the forefront in the study of Naturalistic Decision Making, a new approach in psychology that is directed at understanding the way people use their experience to actually make decisions in operational settings rather than studying college sophomores performing artificial tasks. His contributions have been recognized by the Army, Navy, and Air Force. In 1991, he was the first civilian selected as commencement speaker for the Air Force Institute of Technology. In February 1992, he was invited to give a two-day colloquium on his work to the British Defense Community, at Portsmouth, England.

Dr. Klein received his Ph.D. in experimental psychology from the University of Pittsburgh in 1969. Since then he has conducted numerous research projects on individual and team decision making. The list of projects he has done, articles he has published, and papers he has delivered, runs page after page.

During the mid-1970's, Dr. Klein worked as a research psychologist for the Air Force at Wright-Patterson AFB, and then left to start his own company. The company he founded in 1978, Klein Associates, has grown to a staff of 20 people. In 1989 his company hosted an international conference on naturalistic decision making, funded by the U.S. Army Research Institute, and a book, *Decision Making in Action: Models and Methods*, he and his colleagues have edited will be published in 1993. In the last several years Dr. Klein and his associates have been applying naturalistic models of decision making to a wide variety of topics such as building decision support systems for the Combat Information Center of AEGIS cruisers, for the Weapons Directors of AWACS aircraft, and for the weaponeers trying to do bomb damage assessment with precision guided munitions to reduce the number of unnecessary sorties.

Dr. Klein's work on team decision making has spanned all three services, from Air Force multiship combat in F-15s, to Naval anti-air war operations, to Army projects from the level of platoons up to echelons above corps. The models and methods of group and team decision making are considered by many as original and significant contributions in their own right.

ABBREVIATIONS

ATO	Air Tasking Order
AWACS	Airborne Warning and Control System
BDA	Battle Damage Assessment
CAP	Combat Air Patrol
CO	Commanding Officer
CoA	Course of Action
CRO	Character Read Out
DoD	Department of Defense
DSSs	Decision Support Systems
HCIs	Human Computer Interfaces
IFF	Identify Friend or Foe
JSTARS	Joint Surveillance Target Attack Radar System
NDM	Naturalistic Decision Making
RPD	Recognition-Primed Decision
SA	Situation Assessment
T&E	Test and Evaluation
TADMUS	Tactical Decision Making Under Stress
TQM	Total Quality Management

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

TABLE OF CONTENTS (cont.)

LIST OF FIGURES

LIST OF TABLES

LIST OF EXAMPLES

LIST OF EXAMPLES (cont.)

LIST OF EXAMPLES (cont.)

WHY IS NATURALISTIC DECISION MAKING RELEVANT TO DESIGNERS?

INTRODUCTION

The field of Naturalistic Decision Making (NDM) studies the way people make decisions in operational settings, particularly under difficult conditions. Some of the key models of NDM have emerged only in the last five to ten years, and are just starting to be applied to system design. This State-of-the-Art Report is written to familiarize system developers, human factors engineers, specialists in ergonomics, and design engineers with the ideas and implications of NDM.

The objective of the report is to help you take decision requirements into account for current and future projects. The report attempts to accomplish the following: show you how to identify the critical decision requirements, show you how to use these decision requirements to help organize the system features and the human-computer interface, and help you catch barriers to decision making earlier in the design cycle.

Naturalistic Decision Making has developed methods for defining key decisions, has identified common decision strategies that people use in operational settings, and has described some of the shortcomings of these decision strategies. The intent of this report is to show system developers how to define the key decisions that a new system must support and how to ensure that the system enhances the operators' decision and inference strategies rather than interfering with these strategies.

This chapter discusses the decision-making information that is usually not provided to the designer. The challenge to a NDM approach is to show how decision requirements can be identified and described to support the design process.

Many systems and subsystems, such as Human-Computer Interfaces (HCIs) and Decision Support Systems (DSSs), are built to help users make difficult decisions. The system designers, however, usually aren't given many details concerning the nature of the decisions, or about the strategies used by the operators. There may be information about the products of task analyses, but usually not about the cognitive aspects of the key decisions. Therefore, designers often have trouble figuring out what the system should do, how it should do it, and how information should be presented to the user.

If you are working on a project now, you may find it interesting to look down at Table 1 to see how many of the questions you can answer. The first question is about the key decisions the operators must make, and there may be some clue in the documentation you've been given, or in a task analysis. But those are the official decisions. What are some examples of tough decisions that the operator has to make? For any decision, what specific cues would an operator use? What inferences does an operator have to make?

The point here is that the behavioral technology community has not given you all the tools you need to do your job. Techniques such as task analysis and Data Flow Diagrams trace the observable path of information and control, and they work well for tasks that consist of merely following already existing procedures. They are not so helpful for tasks that require judgment and decision making, because they don't tell you how the operator is thinking through the issues. Data Flow Diagrams portray the way information items are transferred during operations. Figure 1 shows a Data Flow Diagram for the J-STARS self-defense suite. Have you ever tried to trace through a Data Flow Diagram to understand how the operator is wrestling with the task? It's an unfair question; Data Flow Diagrams are not intended to describe the decision task. They are intended to ensure that the necessary items of information will be available, and it is up to you to somehow figure out what form to use in presenting the information.

Table 1.   Questions to ask system developers about the operators'
decision needs.

- What are the key decisions the operators must make?

- What cues do they depend on?

- What relationships between cues are important to monitor?

- How are the operators deriving inferences from the cues?

- Will the new system support these inferences?
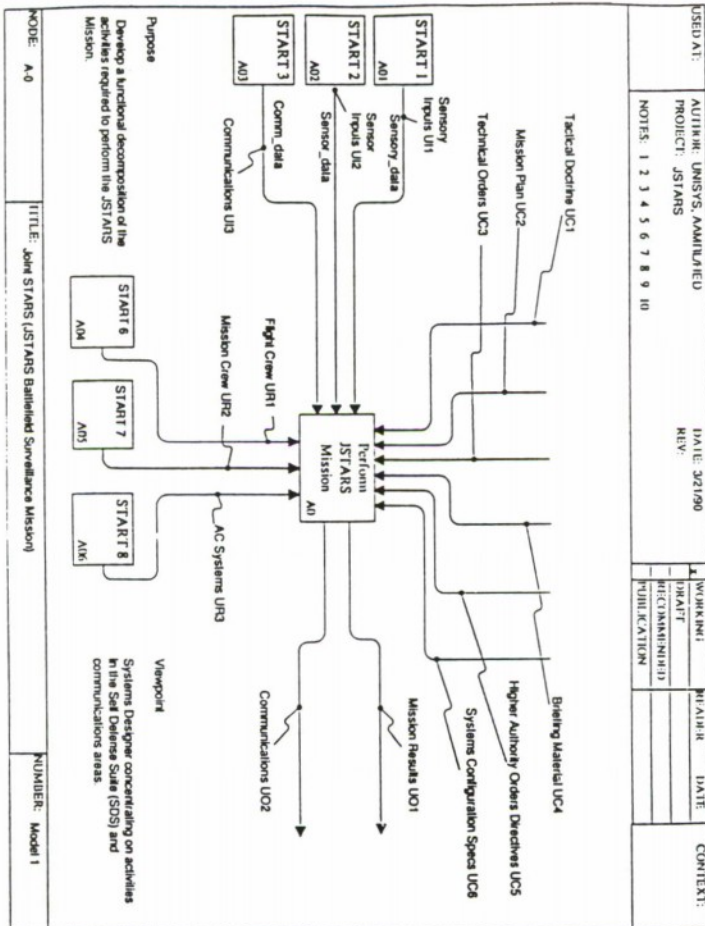
Figure 1. Data Flow Diagram (From *JSTARS--Self Defense Suite Study*. Armstrong Aerospace Medical Research Laboratory, Wright-Patterson AFB, OH, Contract F33615-88-D-0536, November 1990).

Likewise, systems approaches can help you make sure you have identified all the relevant data items, and they can be a good starting point, but they don't show designers what the operator is experiencing.

Consider the last question in Table 1—Will the new system support the important inferences? Even if you have been given a task analysis and a set of Data Flow Diagrams, you can't be sure of your answer, because these only tell you the type of information needed, but not the way it will be used. As the following example shows, just having the right data elements isn't enough if the display doesn't show the key relationships needed to make the inference.

Example 1.1 Critical cues that were missing: Judging fuel flow in a commercial airliner

Once we were observing aircrews handle a malfunction during a simulated flight.[1] We watched three different crews. The malfunction involved a fuel leak. Each time, the flight engineers used their gauges to detect the fuel leak. The fuel-level information had been provided on the fuel gauges. But we also found that none of the flight engineers could estimate the rate of fuel loss with any accuracy. Moreover, when the fuel leak stopped, none of the flight engineers noticed, and the pilots continued to take actions that had become unnecessary. The dials showed fuel levels. The engineers could compare fuel levels for different tanks to detect an abnormal change in one tank and thereby infer a leak. But the dials were not useful for inferring rate of change (the first derivative), or changes in the rate of change (the second derivative). These were critical pieces of information that had been omitted. Simple fuel level was insufficient for managing the emergency. The critical cues for making decisions under time pressure were missing.

The system developers did not deliberately omit information on rates of change. No one had flagged this information as necessary for making judgments and decisions in the midst of an emergency.

If designers don't have a clear picture of the key decisions, or how operators make these decisions, how are they able to build systems and HCIs? Frequently, the best they can do is make sure that a lot of relevant information gets on the screen in a format that appears to be organized. But during actual working conditions the users may find out that the interface is an enemy, not a friend.

Example 1.2   An interface that sometimes got in the way: The control room of a petrochemical plant

Several years ago, during a visit to a control room in a chemical processing plant, a supervisor proudly showed off the new computer-driven system that had just been purchased to replace the old pen-and-ink recorders. "The old system was a nightmare," he said. "The paper was always jamming, and the ink was drying up." But the old system had some important strengths—the operators could look at many different parameters at once. They could check the temperatures at different points in the cycle. Reluctantly, the supervisor admitted that there were many things he could do with the old system, such as rapid re-start, that just weren't possible with the new interface. "We just don't have the same feel for the process with this new set of screens," he said. What bothered him even more was that new operators were starting in with the modern system, and were never going to get a feel for controlling the process.

That is what a NDM framework should do: determine what is needed for skilled operators to have a feel for the process, for being ahead of the curve, for anticipating events. Introducing NDM into the design process should guide the designer in making sure that the system and HCI provide that feel, and even enhance it. By understanding more about NDM, you should learn how to take the

user's needs into account and specify decision requirements, even during early conceptual design when the user cannot articulate these needs because the system is so new.

The idea of taking users' needs into account is a standard one, almost a cliché. What has been missing is a way to elicit the cognitive, decision-making needs and to let the designer understand these decision requirements. That is part of the value added by NDM, to provide direction in figuring out what the user needs.

Two examples are presented where the system developers failed to take the users' decision needs into account. In each example, the system was developed for a newly created type of job, so there were no users to discuss where they were having trouble. The designers needed a way to anticipate the decision requirements.

Example 1.3 Failure to take the user's decision needs into account: Joint STARS

The Joint Surveillance Target Attack Radar System (JSTARS) aircraft is a combined effort by the Air Force and the Army to provide a platform that flies near the forward edge of the battle area, and looks over it to observe enemy troop movements and alert friendly forces as to their makeup and direction. Because the JSTARS aircraft is fairly slow and unmaneuverable, and flies close to the battle lines, it is very vulnerable. Additionally, its radar emissions make it an easy target to acquire. In the initial design for a self-defense suite, the interface for the self-defense suite operator was designed with a menu structure that went several levels deep, and in one place reached down 19 levels.

A complicated menu structure means more time needed to navigate through the system, greater cognitive and memory demands, and thus more potential for error.

Our analysis of decision requirements showed that the self-defense decision was going to be critical to mission success. If the aircraft was too slow to run for protection, it was going to be shot down. If it was too quick to run, it would compromise its mission, because the radar picture of

ground activities took a certain amount of time to establish. Therefore, timing was critical. In this environment, the cumbersome menu design was going to be unacceptable.

For the design engineers who had worked on the self-defense suite, there were no good analogues, e.g., previous systems that were similar and could be used as models.   In some ways, AWACS seemed to be a good analogue case because it also was slow and cumbersome and flew a data-gathering mission at comparable altitudes.   But AWACS was intended to have fighter support in the form of Combat Air Patrols (CAP), and fly much farther back from the battle, because it was concerned with the air picture, not the ground picture. Therefore, AWACS was a useful analogue for some aspects of the mission, but it was a misleading analogue in other ways because JSTARS was intended to fly without dedicated CAP support.

Our decision requirements analysis suggested that the key self-defense decision was going to identify the last possible moment for abandoning course.   This, in turn, suggested several possible HCI and DSS features for monitoring the last possible moment to abandon the mission, depicting lethality ranges for surface-to-air missiles and threatening aircraft, and so on.   The recommended design centered around a graphic display that highlighted the relationship between JSTARS and any threatening aircraft; this design met with strong user approval.

Desert Storm intervened before any enhancements could be made in the self-defense suite. The prototype JSTARS aircraft were rushed into action, and it was necessary to provide dedicated CAP for JSTARS. Had the air war been less successful, the JSTARS aircraft may not have been used because of its limited capacity for self-defense.

Example 1.4 Failure to take the user's decision needs into account: The watchstander station

Several years ago we had a chance to consult with a company that was designing a management information system to help rapid deployment teams react quickly to environmental emergencies. One of the key posts of the rapid deployment team was the watchstander, who would be the first to learn of the emergency and would have to quickly get the word out to the other team members. Because time was so important, and there were so many people to contact, the designers specified an automatic call-out that would send telephone messages to the other team members. To support the watchstander, the designers prepared three screens. The first was a list of the categories of team members (e.g., planners, equipment providers, accounting personnel) because each category would get its own message. The second screen listed the specific people within each of the categories. The third screen showed each of the responders who had been automatically contacted and who had called back to confirm having received the message about the emergency. Figure 2 shows the three screens. So, theoretically, all the necessary information was available on these three screens.

But a key decision was ignored. When the first word came in, and confusion was spreading, what would the watchstander need to decide? The automatic call-out would be taking care of spreading the news. In our analysis, the key decision left for the watchstander was to figure out who had not been contacted yet, and to take the necessary action. If a resource was needed, such as extra telephone lines, and the person responsible for providing it had not called back, then the watchstander would have to find someone else to call. Yet the arrangement of the three screens made that decision almost impossible. The third screen showed who had called back, but the watchstander needed to see who had not yet called back. To figure this out, the watchstander would have to go to screen #1, and for each category, shift to screen #2

Figure 2. Watchstander's station display.

to identify the specific names in the category, and then, holding these in memory, shift to screen #3 to see which of these had called back, and through subtraction, infer who had not called back. All these mental gymnastics would have to be performed during noise and interruptions. In short, this layout of screens was about the perfect way to block the watchstanders from identifying who had not yet been contacted, even though all the pieces of information were available, on the separate screens.

Just presenting the information is not enough. System developers need explanations of how the operators will be making crucial decisions, to achieve a logical and effective design.

This SOAR was written because of frequent concerns that the systems and HCIs being built to help people perform cognitive tasks do not support decision making. Most designers are interested in building systems that have an impact, and they will use ideas from NDM if it helps them to build better systems. One barrier is that design engineers are not given information about decision requirements during concept development and system specification phases.

The field of NDM should do just that—enable users, planners, and engineers to describe decision requirements so that these requirements can guide the design process. Here are the answers to the question, "Why is NDM relevant to system developers?" When system developers are able to take decision requirements into account:

• they will be more likely to build systems that improve performance, particularly for decisions made under the most difficult conditions

• their systems and interfaces will directly support the difficult inferences operators need to make

• they will be able to help the users anticipate decision needs, rather than discover these needs late in the program

• they will avoid the false starts that miss the decision needs of the users, and so will achieve the greater efficiency of a Total Quality Management (TQM) approach

• they will learn how to apply decision-centered design:

-- identify key decisions
-- use these decisions to define system requirements and
to guide system development.

When a NDM approach is used, the decision requirements can
help the designer conceptualize the whole system, rather than just the
decision requirements to the other requirement lists. The following
case illustrates how this might happen.

---

*Example 1.5   The value added by NDM: Battle Damage
Assessment*

*In 1992, after the Persian Gulf War, the U.S. Army
Engineers Waterways Experiment Station contracted with my
company to design a system to aid in assessing weapon
effectiveness in the field. Laboratory and field studies had
developed and tested a new generation of precision-guided,
air-to-ground munitions. The efficacy of these weapons had
become well known to anyone with a television.*

*Weaponeering - choosing the appropriate combination of
munitions, aimpoints, and munition delivery conditions - is very
difficult in the theater of war. Weaponeers in the theater of
war most often do not have engineering backgrounds. The
knowledge and expertise needed to weaponeer hardened
targets is often held by engineers back in the States. Prior
attempts to translate the needed knowledge into a form and
format useful to such users has had mixed results. Further,
weapons are designed and tested in environments where there
is complete knowledge of the target. Weaponeers in the
theater often have incomplete knowledge of the structural
characteristics of targets. Weaponeers also are faced with
other time pressure and workload issues.*

*But the greatest difficulty is encountered after the
weaponeering task, when battle damage must be assessed.
The new generation of munitions does not obliterate targets.
It is difficult to assess the damage caused by a munition that*

penetrates and then explodes inside a structure, leaving an above-ground structure still standing and an underground structure still buried.

The Defense Nuclear Agency, whose mission is to provide technology solutions for the armed services, funded the development of a decision support system to assist in weaponeering and battle damage assessment (BDA). The payoff was clear. If weaponeers in the theater could do a better job of determining what munitions to use and how to use them, and if better BDA could minimize the occurrence of unnecessary restrikes, then for a set of targets fewer sorties would be flown, munitions would be saved, fewer pilots and planes would be placed at risk, and the potential for civilian casualties would be reduced. Theoretically, air-to-ground capacity could be greatly expanded without adding a single new plane, just by improving munition and BDA decision making.

Why were decision specialists brought in to the effort to build a decision support system? Because what might seem to be a straightforward engineering problem was found to be anything but straightforward. It is expensive for a weaponeer to gain field experience: both the munitions and replicas of the specialized hardened targets are expensive. The knowledge gained by experimental results and described by complex mathematical models is difficult to translate to a user with a nontechnical background. Further, the analytical tools used in the experimental and design arenas are not designed for field conditions which require an answer, a prediction of performance, even if data about the target are missing. Even then, the best analysts, with the best data, still might operate too slowly for the pace of war. A potential user in the field stated the challenge: "A 70% answer right now is better than a 100% answer tomorrow."

Building a decision support system made no sense without a prior study of the decision requirements: the nature of the decisions to be made, the information used, the inferences drawn, and the expertise available. My company worked with

the engineers at Waterways Experiment Station applying NDM techniques to identify the available data and their quality, describe the context in which decisions are made, determine the decision and inference strategies that are used, and verify that weaponeering and BDA tools could be used by nontechnical users under operational constraints.

This chapter has explained why decision requirements have a role in system design. The claim is that it is not sufficient to specify the task that the operator must accomplish—the designer needs information about how the operator will perform that task, to give the operator what he or she needs.

The next chapters go into detail about the nature of NDM. Chapter 2 describes what the NDM approach is, and Chapter 3 covers how people make decisions in operational settings. There are two general classes of naturalistic decisions—sizing up the situation, and finding a course of action. Chapter 4 presents some of the most typical strategies for sizing up a situation and making diagnoses; Chapter 5 presents some typical strategies for selecting a course of action. Chapter 6 balances this discussion by examining some of the things that can go wrong during naturalistic decision making—the types of errors that might occur, possible sources of bias, and the impact of stress.

The next few chapters explore ways of applying this account of NDM. Chapter 7 presents an example of how to use NDM for designing interfaces and supports for diagnostic decisions and for action decisions. Chapter 8 explains how to gather the data for deriving decision requirements, and Chapter 9 shows how decision requirements and NDM fit in at various points along the system design cycle. Chapter 10 considers the importance of including team decision requirements in the design process. Chapter 11 concludes with some recommendations and guidelines for using NDM in system design.

WHAT IS NATURALISTIC DECISION MAKING?

Naturalistic Decision Making (NDM) is the study of how people make decisions in the workplace and in their personal lives. Researchers in NDM try to understand and describe the strategies people use in diagnosing a situation or in choosing a course of action. The focus of interest is on actual situations, because decisions are made in context, and the features of context need to be taken into account to understand the decisions.

This chapter describes NDM by discussing the features that mark a domain as naturalistic, tracing the history of decision research into naturalistic domains, and contrasting naturalistic research with laboratory research. The following Chapters 3, 4, and 5 go into detail about the decision strategies found in naturalistic and operational settings.

CHARACTERISTICS OF NATURALISTIC DECISION-MAKING DOMAINS

Researchers in the field of NDM have studied domains that are complex, messy, and challenging. The reason for examining complex domains is to learn how decision makers handle the complexities and the confusion. One of the things that worry us about carefully designed laboratory studies using naive subjects performing well defined tasks is whether the findings will generalize to the real world. Table 2^2 lists ten domain features that are particularly interesting to researchers of NDM. Not every study includes these variables, and

some naturalistic reasoning strategies apply even when most of these features are missing.[3] Nevertheless, the features in Table 2 cover the most challenging aspects of operational settings. To develop systems that help people think clearly under pressure, we must understand how people make decisions under the conditions listed in Table 2.

That is the task NDM research has set for itself.

These characteristics are important for designing systems:

• Most systems must be operated under time pressure. Building a complex menu structure, as in the JSTARS case, makes little sense. If you can anticipate that the operator of the self-defense suite is going to be balancing the vulnerability time (most lethal weapon of nearest threat) versus the security time (minimal time needed to achieve safety, by landing at a friendly base, by using friendly anti-air, or by using CAP), the job of the designer is to help the operator sense this balance.

• Many systems must be operated with ill-defined goals. Again, using the JSTARS example, there is no correct response when an enemy threat appears. It depends on the importance of the mission at that time, feelings of confidence in AWACS and CAP for defense, intelligence about enemy capabilities, skills of the pilot, effectiveness of counter-measures against missiles, and so forth. The system must enable the operator to synthesize all these factors. That's why it wouldn't be helpful to build a decision aid that calculated distances and used an expert system to simply tell the operator what to do.

• Shifting goals refer to the fact that dynamic conditions may change what is important. In the BDA case, mission planners may develop an air tasking order (ATO) that designates certain targets as the highest priority. Under the extremely dynamic conditions of wartime, those priorities may change just hours before mission implementation, with high-priority targets removed from the ATO and replaced by a new set of targets. A support system has to be flexible, to allow the operator to meet changing priorities. Otherwise, the system may be tied up working on an outdated target.

• Data problems are often inescapable. In the BDA case, unreliable or incomplete data made it impossible to use some algorithms. It was important to determine this up front, rather than build the system and later realize that it couldn't be used in the field.

Table 2.   Features of Naturalistic Decision Tasks.

1.   Time pressure

2.   Ill-defined goals

3.   Dynamic conditioning and shifting goals

4.   Inadequate information (missing, ambiguous, erroneous)

5.   Cue learning

6.   Experienced decision makers

7.   Team coordination

8.   Context (higher level goals, stress)

9.   Poorly defined procedures

10.   High stakes

• Decisions are made within the context of larger organizations. An organization will set its priorities and communicate these, sometimes through regulations but more often through evaluations, reviews, and informal emphases. In these ways, an organization sets a culture of practices, prohibitions, and perspectives. For instance, an airline crew responding to a malfunction is reflecting the organization's priorities about safety vs. scheduling. Computer-based decision support systems can alter an organization's communications and its practices, so the organizational impact can be significant.

• Most system operators have some level of proficiency in the tasks they perform, often reaching a high level of skill. Effective systems can take advantage of this experience, but too many systems prevent the operators from using expertise. Expert systems are notorious for interfering with operators' skills. Many of the decision aids built during the 1970s also exhibit this problem. In the Watchstander example, the screens permitted the watchstander to passively monitor the calls coming in, but not to use experience in working around problems and making adjustments.

• Tasks generally involve some amount of teamwork and coordination among different operators. Interfaces and decision support systems can easily disrupt this coordination. In JSTARS, the operator of the self-defense suite needs to make decisions in concert with the mission control coordinator and the pilot. The interface that was designed, however, kept the operators isolated from each other. In a commercial airline cockpit, the flight engineer is junior to the captain, but has access to instruments the captain cannot easily see. During malfunctions, especially nonroutine ones, captains can be put in the position of making decisions without adequate information about the situation, such as the rate of fuel loss.

• Contextual factors such as acute stressors can come into play. Time pressure and uncertainty about data are two stressors. Other acute stressors include noise, high stakes, personal responsibility, visibility of actions, limited resources, and task difficulty. These stressors have predictable effects on decision making, as will be discussed in a later section. For example, the case of the Watchstander illustrates a memory and inference requirement that becomes very difficult during noise and distraction.

 • Operators can't follow carefully defined procedures. Where such procedures exist, there isn't much need for decision making. But even when the procedures have been specified, emergencies and breakdowns can force the operators to invent new procedures on the spot, and to rely on interfaces to permit such flexibility.

 • Finally, the decisions emphasized by the field of NDM involve high stakes, often risk to lives and property. Under these conditions, operators don't need to be motivated. They need to work with systems and interfaces they can trust.

System design can be affected by some or all of these conditions. NDM research is the attempt to learn how to take these characteristics into account. The next section describes how the NDM approach developed.

A BRIEF HISTORY OF NDM

Previous models of decision making avoided the features listed in Table 2. The classical theories of decision making[4] grew out of mathematics and game theory.[5] These models showed how decision makers should use their estimates and judgments to make optimal choices. The models were formulated for straightforward tasks, where decision makers might have trouble synthesizing quantitative data. The models were not intended for cases where time was limited, goals were vague and shifting, data were questionable, and so forth. Therefore, the classical models weren't very useful in designing systems to help people work in dynamic settings. When the classical decision models were used to build decision aids, using Bayesian statistics or Multi-Attribute Utility Theory, the results were usually disappointing, because the users weren't thinking the way the models required. As a result, operators avoided using these decision aids.

NDM is a recent approach. John Payne (1976) and Lee Beach and Terrence Mitchell (1978) pointed out that the classical, heavily analytical decision strategies weren't practical for many tasks, and that under contingencies such as time pressure and uncertainty, people were likely to use simpler strategies. These contingency models still concentrated on how people selected the best course of action (CoA)

from a set of several alternatives. Several years later, Jens Rasmussen (1985) and Joe Wohl (1981) formulated more detailed descriptions of NDM, and linked the functions of diagnosing a situation and selecting a course of action. The previous work has emphasized the task of selecting the best CoA from a set of several. Rasmussen is an engineer and at that time was working to figure out how to build displays in nuclear power plants that would reduce the chances of accidents. Wohl was working on applied contracts for the U.S. Navy to improve command and control. Because neither Rasmussen nor Wohl was a decision researcher, it may have been easier for them to see the relationship between diagnosing a situation and selecting a CoA.

There were a few researchers looking at naturalistic settings. Ken Hammond had studied naturalistic settings for most of his career. For example, he showed that highway engineers made effective use of analytical strategies for tasks such as estimating traffic load, but for other tasks such as estimating accident rates, the engineers did better using intuitive strategies.[6] James Shanteau and Ruth Phelps (1977) found that livestock judges were able to make reliable and accurate decisions without following analytical procedures. Their work stands in sharp contrast to the earlier research that emphasized strategies for selecting one CoA from many.

The critical events for the field of NDM occurred in the late 1980s. Up to that time, there was a growing realization that decision making was more than picking a CoA, that decision strategies had to work in operational contexts, that intuitive or nonanalytical processes must be important, and that situation assessment had to be taken into account. Then, a number of researchers presented models showing how decision makers could use experience to handle operational contexts. Klein (1989; Klein, Calderwood, & Clinton-Cirocco, 1986) reported on fireground commanders and tank platoon leaders and design engineers. Noble, Boehm-Davis, and Grosz (1986) reported on Naval command-and-control personnel. Pennington and Hastie (1981) reported on jurors. Beach (1990; Beach & Mitchell, 1978) studied business decisions. Lipshitz (1989) reported work with Army officers. (See Klein, Orasanu, Calderwood, & Zsambok, 1993, for more detailed accounts of all this work.) It is one thing to point out the

limitations in classical models of how optimal decisions should be made. It is another to formulate models of how decisions actually are made in operational settings. With the emergence of these models, NDM research achieved coherence as an approach for studying basic and applied issues.

CONTRAST BETWEEN NATURALISTIC DECISION MAKING AND LABORATORY RESEARCH APPROACHES

Some people have argued that any setting is naturalistic, including laboratories to study college students. Certainly, controlled laboratory studies have discovered some useful things. One premise of NDM is that there will be a higher payback for studying decision making in more realistic settings. For one thing, there are some variables, such as personal risk, that cannot be studied in the laboratory. Also, laboratory studies cannot re-create many other conditions listed in Table 2, so applied researchers are usually reluctant to generalize from these studies. Finally, carefully controlled laboratory studies have difficulty in meeting the criteria of (a) trying to understand the strategies people actually use and (b) occurring in settings that contain most of the features listed in Table 2.

Many laboratory studies of decision making are based on quantitative models in subject areas like economics, statistics, probability, and game theory.[7] These models lent themselves to carefully controlled research. Now that we have learned the boundaries of these models and their limitations for handling the features in Table 2, we want to describe how decision makers actually function. The NDM approach is attempting to provide such a description, and to do that it has been necessary to observe the phenomenon of decision making as it actually occurs. These observations continue to be refined into models and hypotheses. The NDM approach is to use field studies, interviews, observations, and realistic simulations of tasks to identify the processes and variables of importance. Too often, laboratory research has relied on artificial problems, limited context, and naive subjects, and as a result has limited its own progress. We hope to learn how to conduct laboratory

studies that incorporate sufficient realism to make the results generalizable.[8]

The features in Table 2 are directly relevant to professionals interested in system development, particularly systems that will support the operators' decision making. The NDM framework attempts to clarify the strategies people use to make decisions in domains marked by these features.

In this chapter we have seen the features of field settings that designers anticipate, and we have traced the growing interest in models of decision making that apply to field settings. The next few chapters describe the models and strategies of naturalistic decision making.

3

HOW PEOPLE MAKE DECISIONS IN
NATURALISTIC SETTINGS

To provide an overview, this chapter will describe the general strategies people use in naturalistic settings. The following Chapters, 4 and 5, provide a more specific look at strategies for making diagnoses and selecting courses of action.

The most important finding that emerged from NDM research[9] was that, in actual cases, people rarely compared any options at all. For example, Klein, Calderwood, and Clinton-Cirocco (1986) tried to find out how fireground commanders made decisions about how to deploy their crew members during the most difficult urban fires they had faced, but the commanders insisted that they never tried to figure out whether one option was better than another. For researchers trained to expect that decision making necessarily involved comparison between options, this was totally unexpected. How can skilled decision makers select effective courses of action without comparing options?

SELECTING ACTIONS WITHOUT COMPARING OPTIONS

Research in NDM has shown that decision makers can use their experience to size up the situation, recognize it as typical in some ways, and identify the typical way of responding. Therefore, skilled decision makers may never have to consider more than one option. The different strategies for contrasting options[10] rarely come into

play. Of course, there are times when it is important to contrast options, and this will be discussed later. But for most cases, including very difficult incidents, the critical step is to assess the situation.

This is illustrated by an incident described by Kaempf, Wolf, Thordsen, and Klein (1992), in which the commanding officer of an AEGIS cruiser needed to decide whether to shoot down some threatening F-4 airplanes. On the surface, the decision was about different CoAs—firing missiles or not. On a deeper level, the decision was about assessing the intent of the F-4 pilots. A decision flow diagram for this incident is shown in Figure 3.

Example 3.1 NDM: The Harassing F-4s

In 1988, the Iran-Iraq war had endangered shipping in the Persian Gulf. An AEGIS cruiser was patrolling the Persian Gulf, to keep the sea lanes safe. On this particular day, the cruiser was escorting its unarmed flagship through the Gulf, in daytime. Two Iranian F-4s took off and, instead of patrolling the coast to the north or south, began to circle the end of the runway. Each orbit brought the fighters closer to the U.S. Navy ships. The aircraft turned on their search radars, to scan for objects. Then the lead aircraft turned on his fire control radar used to obtain a radar lock-on to a target prior to firing a missile and acquired either the AEGIS cruiser or the flagship as a target. This was considered a hostile act, and the commander would have been justified in firing a missile at the F-4s. However, his mission was to reduce hostilities, not increase them. He needed to defend his ship, and the flagship, but in his judgment the F-4s were not going to attack.

He formed his judgment by trying to imagine that the F-4s were hostile. He could not imagine that a pilot preparing to attack would make himself so conspicuous. The pilots had been flying around in plain view. They further announced their presence by turning on their radars. They even used their radars unnecessarily, keeping them on when their circles carried them away from the cruiser. This was particularly

unusual because the Iranians were having trouble performing maintenance on the radar systems, and tried to use them as little as possible. Yet here were aircraft making a big show of using their radars. The commander just didn't see how pilots intending to attack him would behave that way.

In contrast, he could imagine how the pilots were trying to harass him. All their actions seemed consistent with the harassment hypothesis, whereas the hostile intent hypothesis had some major flaws. Therefore, the commander inferred that the F-4s were just playing games.

He still needed to ensure self-defense, and he took the necessary actions — breaking the lock-ons from the F-4 radars, sending out radio warnings, and so forth. He also prepared his crew to look for telltale signs, such as swerving away, that might indicate that the F-4s had fired missiles. Finally, he determined the minimum range he could accept, and prepared to engage the F-4s if they got too close. Eventually the F-4s tired of the game, and flew off.

In their review of this incident, Kaempf et al. note that the core of the decision was inferring the intent of the F-4s, and that once this was done, the commander knew how to make the decision about firing missiles. The commander was not going to be a victim of tunnel vision, because he was aware his diagnosis might be wrong, and he prepared his crew to take immediate action if the fighter aircraft got any closer.

An analytical approach might have framed this as a shoot/no shoot decision, assigning probabilities to the hypotheses about whether the F-4s were hostile or harassing, estimating utilities for shooting if hostile, shooting if harassing, not shooting if hostile, and not shooting if harassing, as in Figure 4.[11] Decision aids were built to help perform these types of decision analyses. From the viewpoint of the NDM paradigm, Figure 4 is irrelevant. It does not describe what the commander was trying to do, which was to use every bit of information to build a plausible story of what the F-4 pilots had in

NDM - Implications for Design
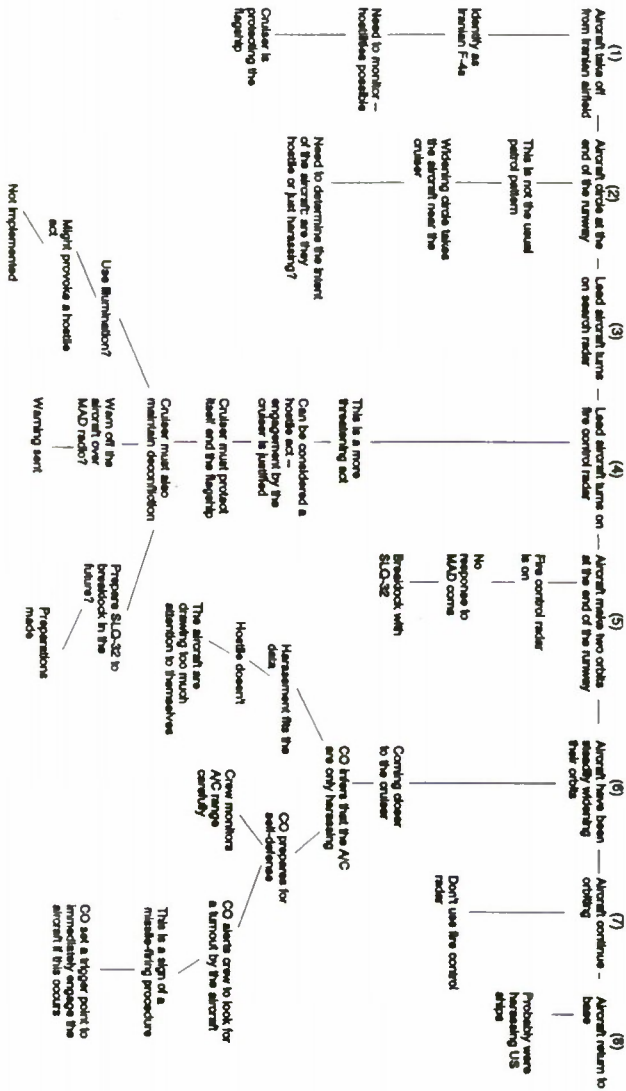
Figure 3. The Harassing F-4s: Decision flow diagram.

mind. Furthermore, encouraging decision makers to go through a process as shown in Figure 4 is not very helpful, and often gets in the way. When a NDM framework is used to design a decision support system and interface, the results look very different, as will be described in Chapter 9.

In this incident, you can see that the commander was able to infer the intent of the F-4 pilots, using his experience about how a fighter pilot would conduct an attack. Once he had inferred the intent, the CoAs were fairly obvious. The incident illustrates a key set of insights from NDM research, presented in Table 3.

The first claim is that most of the time people try to satisfice and find a workable solution, rather than optimize or find the best solution. Simon (1955) was the first to make this distinction, based on his observations of business decisions. In operational settings it is very hard to figure out what the best CoA is, even with hindsight. Decision strategies that try to calculate the best CoA only work when time is plentiful and the goals are very clearly defined. In the incident with the harassing F-4s, no one can say that the commander was right or wrong in not firing missiles as soon as the F-4s used their fire control radars. In this case it worked out, because he avoided an incident by increasing his level of risk while retaining his ability to defend his ship.

The second claim is that we have to distinguish situation assessment decisions from CoA decisions. Sometimes, we need to diagnose what is going on and, perhaps, to select one diagnosis from several possibilities. Other times, we need to figure out which action to take. In the F-4 example, the commander was faced with a diagnosis decision.

The third claim is that, in operational settings, people can use experience to arrive at a situation assessment. They can use context (such as knowledge of problems with radar maintenance) to help them draw inferences. They can trust this ability to arrive at a situation assessment.

The fourth claim is that, in most cases, the situation assessment makes it obvious how to respond. In the F-4 example, once the commander inferred that the F-4s were just harassing him, it was not
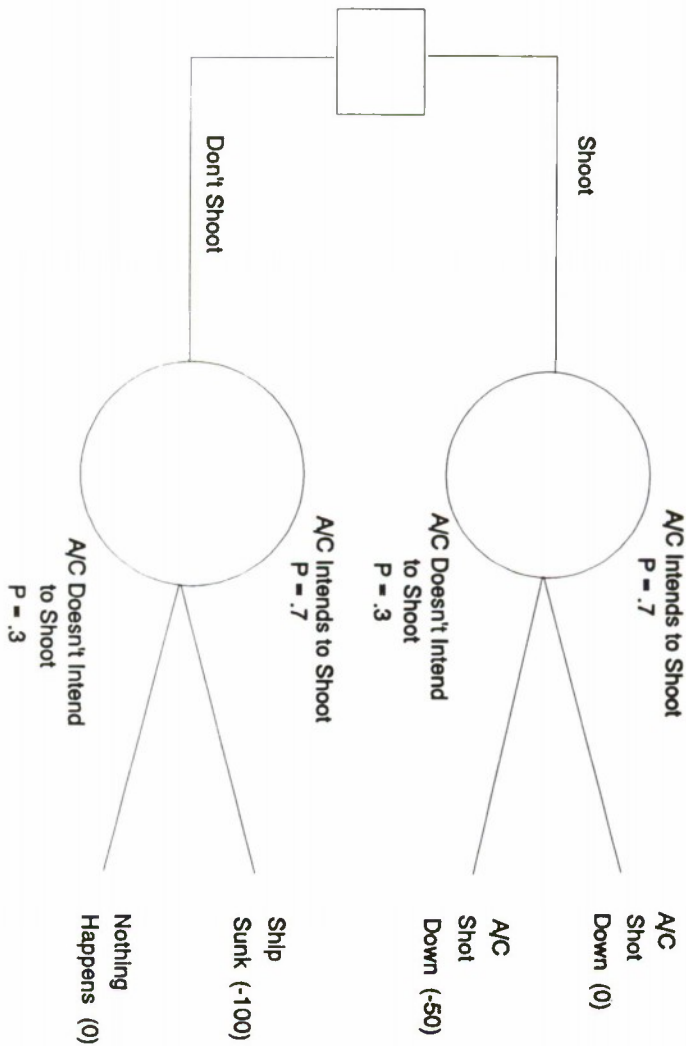
Figure 4.   Decision analytic representations of F-4 incident.

Table 3.   Seven Claims of Naturalistic Decision Making.

1.   In operational settings, people try to find the first course of action
     (CoA) that works, not the best one.

2.   Decision making consists of two aspects--assessing the situation,
     and selecting a CoA.

3.   Experienced decision makers can usually assess the situation
     quickly and accurately.

4.   Once the situation is understood, the CoA decision is usually
     obvious.

5.   Decision makers often must be prepared to act without fully
     examining the parameters and contingencies.

6.   Decision making and problem solving are inter-related.

7.   Decision makers arrive at a CoA by generating pertinent options
     rather than filtering out unacceptable options.

reasonable to think about firing missiles at them, as long as their range did not make them too much of a threat. There are standard ways of responding to unwanted harassment, such as radio calls and breaking lock, and the commander took these steps.

The fifth claim is that decision makers usually must act without having all the facts. In the F-4 example, the commander didn't know for sure what the Iranian pilots were up to. He didn't know what weapons they were carrying. While he assumed the best case about their intentions, he also had to assume the worst case about their weapons and prepare to fire at them if they got too close.

The sixth claim is that NDM is tied to the field of problem solving. It was easy to keep these two topics distinct as long as classical decision-making research (see Baron, 1988; von Winterfeldt & Edwards, 1986, for reviews) focused on game theory, probability estimation, and statistical analyses. However, in a naturalistic context, the problem-solving requirements of clarifying goals and evaluating possible solutions must merge with the decision requirements to assess situations and evaluate possible solutions. Klein and Weitzenfeld (1978) have pointed out that most naturalistic tasks involve ill-defined goals, so that the usual advice to first define the goal and then search for options is inappropriate, because the process would never get past this first step. Instead, Klein and Weitzenfeld assert that for ill-defined goals, we press on and attempt to find solutions. When we evaluate inadequate solutions, we learn about new goal properties and improve our understanding of the goals. By simultaneously searching for solutions and increasing goal clarity, we are enabled to solve the problem. Duncker (1935/1945) was the first to show how goal seeking and goal clarification were interrelated. Early work in Artificial Intelligence (Newell & Simon, 1972) sought to build on the work of researchers such as Duncker, but became directed at well-defined goals because these were amenable to demonstrations. In the F-4 example, we can say that the commander was making a decision about the intent of the airplanes, and treat it as a diagnostic decision; or we can say that the commander had to figure out how to keep at arm's length two airplanes that were probably not going to attack, and treat it as a problem to be solved.

The seventh claim is that decision makers identify and select a CoA based on option generation and problem-solving activities. Option generation sometimes relies on memory for an analogue experience, or memory for a prototypical case, or creative construction to adapt and strengthen options.[12] It is sometimes useful to consider more than one option to identify important requirements and distinctions, but the intent here is usually goal clarification. This approach stands in contrast to the view that decision making is a largely negative process in which inadequate CoAs are filtered out. The decision maker starts out with a large set of options, eliminates most of these by testing for simple features, winds up with a small set of finalists, and rejects all but one. Standard advice[13] is to generate a large set of options, to make sure no good CoAs are missed, and then carefully screen out the worst ones. This may seem like wise advice, but it is time consuming and memory intensive. The NDM framework claims that decision makers don't follow this advice, and that they don't use systems that are based on screening and filtering.

What is new about NDM? These seven claims already constitute an important departure from classical decision models. They portray decision makers as capable of using experience to handle difficult situations, without having to evaluate different options. One of the comments people make after hearing about NDM is that it all sounds obvious—surely people knew all that already. In fact, prior to the NDM approach, the standard advice you probably received about how to make better decisions was to generate many different options and carefully list the strengths and weaknesses of each, to calculate the best. Anything less than a careful generation of sets of options, a careful preparation of evaluation dimensions, and a careful assignment of values, was seen as deficient.[14] The seven claims listed above raise serious questions about standard decision training. According to the NDM framework, this advice to generate and contrast different options may be useful to novices, who lack an experience base for diagnosing situations. But the advice is incompatible with the way proficient operators make decisions. The available data[15] clearly show that decision makers do not follow the classical advice of contrasting options. Furthermore, departing from the classical advice is what experts are able to do and is not a cause for guilt. Clearly

4there are times to compare options, particularly for novices. Chapter 5 examines the conditions under which it makes sense to contrast options.

Example 3.2 NDM: The Recognition-Primed Decision (RPD) model

The RPD model[16] was developed to describe how people can make good decisions without ever comparing options. The initial studies were done with fireground commanders. We expected that they would use their experience to cut down the number of options they compared, maybe just looking at two. We were wrong—they insisted that they hardly ever compared options. In our interviews with them about how they made tough decisions, we kept hearing about the same type of strategy. We derived the RPD model from what they told us.

There are two components of the model, situation assessment and option evaluation. The RPD model asserts that decision makers recognize the dynamics of a situation, enabling them to identify a reasonable course of action, and this CoA is evaluated by imagining how it will be implemented. Experienced fireground commanders can size up a fire pretty quickly. By assessing the type of fire and the type of structure, it usually is obvious how to respond. Still, the stakes are high, and errors can be costly, if not fatal. How do you evaluate a course of action if there are no others to compare it with? One strategy fireground commanders use is to imagine carrying out the action. They run it through in their minds. Sometimes they run it through several times, if the risks are very great, or if the course of action is complex. We have called this process "mental simulation," because they are simulating the course of action in their heads, to see if it will work. This process of mental simulation—mentally enacting a sequence of events—can appear also in the situation assessment phase of the model.

Figure 5 shows two versions of the RPD model. The simple version appears in the panel on the left, where the decision maker confidently identifies a situation as familiar.

When you recognize the dynamics of a situation, you know several important things.

- *You know what goals make sense, so you don't waste your energy on foolish schemes.*

- *You know what cues are relevant, so you don't get overwhelmed by all the information.*

- *You know what to expect so you can be prepared, and also so you can notice surprises, which may mean your diagnosis was wrong.*

- *You know the typical ways of reacting, so you are poised to respond when necessary.*

The panel on the right shows a more complex RPD strategy. Here, the situation assessment was not so easy. The decision maker may have needed to acquire more information. Or else, there were several different hypotheses about what was going on. For instance, in the F-4 example presented on p.24, the commander had to choose between two different hypotheses. Either the aircraft were harassing him, or else they were preparing to attack him. Decision makers use different strategies to arrive at a situation assessment, or to choose among different situation assessments. One is to check the hypotheses against the features of the situation. The other is to build a mental simulation, or story, to explain the events. In the F-4 incident, the commander tried out one mental simulation, that the planes were preparing to attack him, and judged that it didn't make sense. He couldn't construct a plausible story of how a pilot would make such an attack. The other story did make sense, and he went with it.

Once the decision maker settles on an interpretation of the events, the same functions are accomplished as in the simple
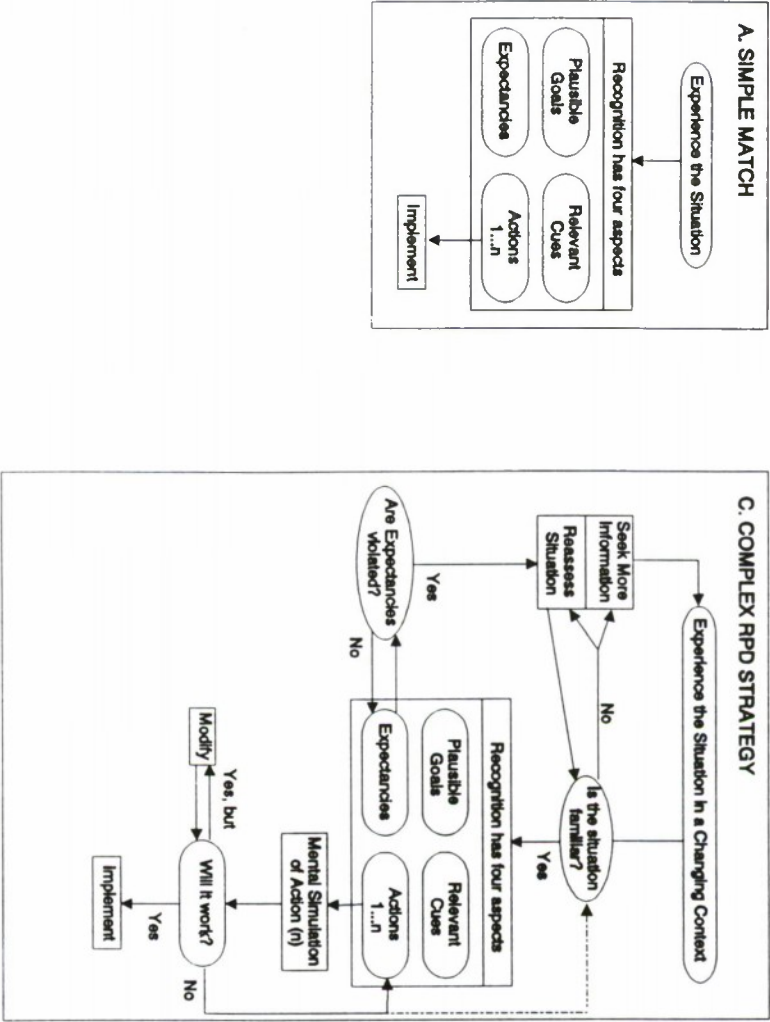
Figure 5.   Recognition-Primed Decision model.

RPD model: specifying plausible goals, highlighting critical cues, generating expectancies, and identifying reasonable courses of action.

In complex cases, the expectancies can be violated, leading the decision maker to seek more information and to reassess the situation. Complex cases can also call for evaluation of a CoA. From the interview data we have collected, there seem to be two primary ways of evaluating options: checking them for necessary features, and using mental simulation. Sometimes people just evaluate a CoA by checking to see if it has the required features, and don't go through any mental simulation at all. For example, decision makers may reject one option after another, until they find one that works. They may consider a number of different CoAs, without ever comparing one to another. That is, they evaluate the options one at a time, until they find one that works. This is called a singular generation/evaluation process, to distinguish it from settings where people are trying to compare different options to each other.

In more complex cases, once a reasonable CoA is identified, the decision maker may try to imagine how it will work in context. This is still a singular generation/evaluation of options. If you are concerned that F-4s may be preparing to attack your ship, one obvious response is to fire chaff, to distract an enemy missile. But if you play this out in your head, you may realize that your ship is between the Iranian fighters and the flagship you are defending, so if you fire chaff it may divert the missiles away from you and directly towards your flagship. In this case, the Electronic Warfare coordinator mentally simulated the problem and rejected the option of using chaff. In other cases, mental simulation helps to show problems that can be overcome, so that the course of action is strengthened.

Since its development, the RPD model has been verified in domains other than firefighting. It describes the decision strategies used by tank platoon leaders,[17] commanders and Anti-Air Warfare officers of AEGIS cruisers,[18] nurses in

Intensive Care Units,[19] commercial pilots,[20] and even design engineers![21]   We have tested the model using simulated firefighting incidents.   Furthermore, the RPD model predicts that experienced decision makers can generate a plausible CoA as the first one they consider, and we verified this hypothesis in a study of chess players.[22]

The value of the RPD model is to:

• explain how people can use experience to make decisions

• describe how decision makers can use situation assessment to identify a CoA

• describe how decision makers can settle on a CoA without considering any others

• show how people using mental simulation can strengthen a CoA rather than choosing only from the set of original options

• describe how decision makers can be poised to act, rather than having to wait to complete their comparisons and analyses

---

There are a number of descriptions of the NDM approach in addition to the RPD model. Hammond, Hamm, Grassia, and Pearson (1987), Beach (1990), Cohen (1989), Connolly and Wagner (1988), Lipshitz (1989); Montgomery (1983), Noble (1989), and Pennington and Hastie (1988) have presented important contributions, some of which will be discussed below.[23]

Seven claims of the NDM paradigm were in Table 3. We can go deeper than these seven claims. The NDM framework has more specific assertions to make about situation assessment and CoA decisions. These are covered in the next chapters. Chapter 4 addresses the strategies used to form situation assessment and make diagnoses, and Chapter 5 covers the decision strategies for selecting a CoA.

4

DIAGNOSIS AND SITUATION ASSESSMENT

According to the NDM framework, situation assessment (including diagnosis) is the most important aspect of decision making. In designing systems, providing decision support, and building interfaces, it is essential to help the operator understand what is going on and how different variables are interacting. System and interface requirements for helping a person form a situation assessment differ from those for helping to choose a CoA. Decision aids that de-emphasize situation assessment and highlight the CoA component can actually interfere with performance. That is why it is important to understand the role of situation assessment in decision making. The NDM framework can help us here by clarifying what situation assessment involves.[24] First we will examine the primary aspects of situation assessment. Then we will enumerate some of the functions that situation assessment provides. Next we will describe some of the common strategies for arriving at a situation assessment. Finally, we will touch on the contents of situation assessment.

FOUR ASPECTS OF SITUATION ASSESSMENT

A decision maker who understands the dynamics of a situation knows four things: feasible goals, relevant cues, expectancies, and plausible CoAs. (See Figure 5.) These hold whether the recognition is immediate, so that the task is judged as familiar, or whether the situation assessment requires conscious inference. A decision maker

with an adequate situation assessment is:
- pursuing appropriate goals
- noticing relevant cues
- confirming events as they occur
- preparing to carry out reasonable actions

In the example involving the F-4s, the commander concluded that the fighters were "gaming" rather than posing a severe threat. Therefore, one goal was to avoid escalating the conflict; so he didn't use his radar to lock on to the aircraft, for fear of startling or provoking them. Another goal was to send out periodic signals that he was monitoring the F-4s, using radio warnings. He also communicated with the F-4s by breaking lock every time they came around, and this satisfied his goal of keeping up his defenses. Knowledge of goals is important, because different functions and cues become relevant as goals shift.[25]

The commander's understanding of the intent of the Iranian jets conditioned him to keep track of certain cues, such as range, altitude, speed, use of radar, radio responses, and to ignore other cues such as friendly tracks on other parts of the screen, identify friend or foe (IFF) signals which were irrelevant once the identification had been made. Good interfaces make it easier to find the relevant cues on the display.

Expectancies are the sign of an experienced decision maker, who has seen similar events and knows the different ways that events play out. In our example, the commander assumed the F-4s would eventually leave him alone. Nevertheless, the stakes were sufficiently high for him to anticipate what might happen if he were mistaken. He alerted his crew to these violations, such as a sudden turning out by the airplanes, which is found after missile release, as the plane guides the missile towards its target while heading towards safety.

A decision maker who recognizes a situation also recognizes the typical reactions that are possible. This is the basic principle of accounts such as the RPD model that explain how operators can translate experience into action.[26] Klein and Crandall (1992) have suggested a convergence model of option generation, in which the situation assessment of feasible goals and patterns of resources together generate the options that are considered.

Functions Provided by Situation Assessment

By providing the decision maker with the four types of knowledge discussed above, situation assessment provides the basis for a number of important functions. A clear situation assessment lets the decision maker:
- use expectancies to verify if the situation assessment is accurate
- use expectancies to guide behavior
- identify a favored CoA
- monitor a CoA to detect problems and diagnose them
- manage resources such as attention, information seeking, and time
- prioritize actions to reflect goals and constraints
- develop plans

In designing and evaluating systems, these are the criteria to use in determining whether a system concept will do its job.

Strategies for Developing Situation Assessment

One of the most difficult tasks is to diagnose a problem, sifting among the symptoms and clues to infer what is happening. The example of the F-4s was centered around diagnosis—inferring the intent of the pilots, using their different behaviors. Diagnosis is central to medical decision making and to troubleshooting, as well as to a variety of other fields and tasks. Despite its importance, there has been relatively little work on the strategies people use to make diagnoses in naturalistic settings. Much of the work by researchers studying decision making and problem solving has used well defined tasks, with limited context, to see if subjects could accurately estimate the likelihood of different hypotheses given probabilistic evidence. This is the rationale for studies of Bayesian statistics.

Our interest is in the strategies people actually use in diagnosing conditions where data are missing and ambiguous, parameters keep changing, actions affect the cues, and relationships are complex and uncertain. Consider an accident in a nuclear power plant. The effects may occur some time after the causes. There may or may not be

multiple faults. Attempts to reduce some of the symptoms have their own effects, sometimes unpredictable effects, so the controller cannot be sure if the apparent symptoms are real, or artifacts of prior control actions. If the controller could determine the true state of the plant, it might be obvious what to do. The difficulty is in untangling the different signals and making sense of them.

Research on situation assessment is fairly recent, so the list of identified strategies is brief. We expect this to change in the coming years.   There are three primary strategies to consider: feature matching, analogical reasoning, and mental simulation.

The most frequent strategy is for a decision maker to use a combination of feature matching and pattern matching, to judge that the events are so close to a given hypothesis that the hypothesis can be adopted as the explanation. Often, this judgment is made without awareness. But there are times when the decision maker deliberately reviews the features to see where they match and where they miss, and whether the mismatch is critical or can be argued away. Feature matching also becomes a deliberate strategy when there are alternate hypotheses and the task is to determine which fits the data better. Noble (1989) has been the researcher most responsible for showing that feature matching can be used for situation assessment in NDM settings.   Noble's work is particularly important because he has developed generic software for building feature-matching decision support systems, to alert operators when features match pre-defined hypotheses.

Reasoning by analogy is another important strategy. Sometimes, a decision maker will retrieve and use an analogy without thinking about it, and at other times there will be a deliberate search for an analogue, and a careful mapping of the analogue onto the situation. Klein and Weitzenfeld (1978) have described the importance of analogical reasoning for diagnosing a problem, and there have been a number of key research projects[27] to clarify different aspects of analogical reasoning.   Recently, analogical inference has become relevant to system designers because of the interest in Case-Based Reasoning, a computational approach to expert systems that uses analogues rather than rules.[28]

One of the most interesting strategies for diagnosing a situation is

mental simulation. A decision maker may try to imagine a sequence of events that would explain the pattern of cues that are observed. Tversky and Kahneman (1974) were the first to point out the importance of mental simulation as a heuristic.[29] In their description of mental simulation, Klein and Crandall (in press) suggest that a decision maker can use an initial state, and build the simulation forward in time (as in evaluating a CoA), or work from a current state and build the simulation backwards, to figure out what the starting point must have been; this is how diagnosis occurs.

Mental simulation is a source of power for decision tasks such as diagnosis. In a nuclear power plant accident, the troubleshooters would need to imagine different faults and their propagation. In the AEGIS cruiser example, the commander tried to imagine a pilot with hostile intent showing the observed pattern of behavior. In essence, this amounts to building a story. Pennington and Hastie (1988) have documented how jurors rely on story building to formulate hypotheses about what happened during the incident being tried. Pennington and Hastie also suggest criteria for evaluating stories, such as consistency and plausibility. Beach (1990) has put forward Image Theory, which covers a full range of decision functions, and posits the use of mental simulation to define goals (a part of situation assessment), and to imagine how a situation will develop if left alone. Elstein, Shulman, and Sprafka (1978) have performed an extensive study of medical diagnosis. They found that physicians are taught to suspend judgment until all tests have been conducted, but in actuality physicians are quick to generate hypotheses, and they use these hypotheses to suggest the tests to conduct. These hypotheses enable the physicians to imagine how a condition or disease evolved over time, to produce the set of symptoms reported. Hypotheses would be context bound, because the same disease might develop differently depending on the age, size, and health of the patient. This use of hypotheses seems related to mental simulation.

As we learn how to support mental simulation, using different types of aids and display techniques, we will be able to make important progress in system design. Many displays present various components needed to build a story or assemble a mental simulation, but there is much more that can be done to help the operator piece

together a coherent account of events. Some concepts for helping operators use mental simulation for diagnosis will be described later, in Chapter 7.

Situation assessment includes more than diagnosis. The example of the self-defense suite in JSTARS (Example 1.3) illustrates situation assessment without diagnosis. The operator did not have to try to infer underlying causes and dynamics. The basic situation assessment judgment was made when the JSTARS aircraft was about to face unacceptable risks. The operator might need to calculate at what point JSTARS would have to break off its mission because a threatening aircraft had gotten too close. The strategies of feature matching, analogical reasoning, and mental simulation seem to apply to this type of judgment as well as to diagnostic decisions. But there is a difference between forming a situation assessment to reflect the different parameters, as in JSTARS, and a diagnostic decision to imagine why certain events have happened, as in Example 3.1 of the F-4s that were thought to be harassing.

THE CONTENTS OF SITUATION ASSESSMENT

In many settings, situation assessment is used to refer to what a person knows—the content of knowledge, not its form. For instance, pilots who maintain awareness of events during air-to-air combat are said to have good situation assessment. When they get confused and make a mistake, they are said to have poor situation assessment. The content of knowledge is obviously important. For pilots, that knowledge would include their own speed, the adversary's speed, relative altitudes, headings, bearings, weapons status, fuel status, position of other friendly and enemy aircraft, anti-air missile sites, and so forth. A good pilot needs to take all this information into account, plus a whole lot more. And this is just the beginning. There are many interrelationships and secondary cues to consider as well.

One difficulty here is that the concept of situation assessment loses its meaning. It simply designates all the important things to which a pilot needs to attend. If a pilot makes a mistake, in hindsight it is easy to say that it was because of poor situation assessment, but there must

be many occasions in which a pilot got confused about a parameter and recovered, with no one accusing him of losing situation assessment. In this usage, situation assessment becomes an after-the-fact scapegoat for errors, a general purpose excuse.

For system designers, it is important to identify all the critical cues and relationships that go into situation assessment, to make sure that the operator gets all the necessary information. But specifying critical cues is not enough. The designer also needs to understand the strategies on which the user will rely to build diagnoses, and the functions that situation assessment serves, to specify systems and interfaces that are clear and easy to use, even under time pressure and other stressors. The content items are the starting point. The designer needs to visualize how the operator will use the critical cues, to come up with a coherent and integrated system specification.

This chapter has reviewed the concept of situation assessment, describing the knowledge provided when a person achieves situation assessment, the various functions that situation assessment supports, the strategies people use in diagnosing a situation and forming a situation assessment, and the contents of situation assessment.

In building interfaces that improve decision making, perhaps the most effective approach is to improve situation assessment. This chapter attempted to provide ideas about where an effective interface could support situation assessment.

SELECTING A COURSE OF ACTION

Naturalistic Decision Making research is concerned with understanding the strategies people actually use to arrive at a CoA. Emphasis is placed on situation assessment because that drives the CoA decision. But the bottom line is to explain how people faced with time pressure, uncertainty, missing data, and unclear goals, can select a CoA to carry out. The system, including support features and interface, must enable the operator to react quickly and effectively. The NDM paradigm can help us here by clarifying what a CoA decision involves. First, we will examine the primary aspects of a CoA decision. Then we will enumerate some of the functions that a CoA decision provides. Finally, we will describe some of the common strategies for arriving at a CoA decision.

Table 4 contrasts the different aspects of situation assessment decisions and CoA decisions. It summarizes the discussion of situation assessment and shows the linkages and contrasts between situation assessment and CoA decisions.

Each type of decision generates certain types of knowledge as an output. Each type of decision enables a person to perform certain types of functions. And each type of decision relies on characteristic strategies. The strategies used for situation assessment are also used for a CoA decision made using singular evaluation; comparative evaluation depends on a different set of strategies.

To understand the relationship between situation assessment and CoA decisions, it may be helpful to use a chemical analogy.

Consider a water molecule. It consists of two hydrogen atoms and an oxygen atom. The properties of water emerge at the level of the

Table 4.   Comparison of SA and CoA decisions.

	Situation Assessment	Course of Action
Outputs:	Identify feasible goals	Identify a workable option
	Define relevant cues	Understand +/- of the option
	Generate expectancies	Appreciate counter-indicators
	Identify a workable option	
Functions:	Verify SA accuracy	Resolve the incident
	Guide behavior	Prepare to resolve the incident
	Identify a favored CoA	Manage resources
	Monitor a CoA	
	Manage resources	
	Prioritize actions	
	Develop plans	
Strategies:		Singular Evaluation:
	Feature matching	Feature matching
	Analogical reasoning	Analogical reasoning
	Mental simulation	Mental simulation
		Comparative Evaluation:
		Atomistic strategies
		Global strategies

molecule, and cannot be predicted from knowledge of the atoms alone. Yet we know better how to use water in complex reactions if we understand the atoms that compose it. Similarly, naturalistic decisions usually involve situation assessment and CoA, and we don't fully understand NDM if we only pay attention to the CoA aspect. Models of NDM, such as the Recognition-Primed Decision model, incorporate both situation assessment and CoA. In fact, the RPD model includes all three of the common strategies in Table 4, feature matching, analogical reasoning, and mental simulation. The RPD model does not cover comparative evaluation, because the intent of the RPD model is to explain how people can arrive at a CoA without contrasting different options.

For a system developer, if you want to go beyond knowing the steps a task requires and to gain a sense of the way the operator performs these steps, Table 4 should help you see the differences and interrelatedness of situation assessment and CoA decisions.

THREE ASPECTS OF SELECTING A COURSE OF ACTION

We can distinguish three types of knowledge that are required when a decision maker chooses a course of action. These aspects are listed in Table 4. They are to identify the option, to understand its strengths and weaknesses, and to appreciate indicators that can serve as warnings that the option is not working out well.

According to the NDM framework, in most cases the situation assessment will identify a workable option. There are also times when the option will come from other sources, such as a supervisor, a colleague, or even a mechanical set of rules that were drawn up to cover the task. The CoA may even be suggested by a decision support system. Whatever the source, identifying one or more options is essential to decision making. But it is not always sufficient.

Depending on how the option is identified, the decision maker may understand its strengths and weaknesses. If the option comes from recognizing the situation as typical, then a person may draw on previous experiences or folklore to anticipate what the option is offering, and where its shortcomings are. In contrast, if the option is

suggested by someone else, or by a decision support system, then this background information is likely to be missing. If decision makers don't have a good understanding of strengths and weaknesses, it can be risky to implement the CoA.   Perhaps there is a decision requirement for support systems to provide background information along with recommendations.

A third type of knowledge associated with a CoA is knowing when the CoA is failing. With experience, we learn to recognize the early signs that a plan is falling apart. We also can develop contingency plans, when necessary. It is easier to commit to a CoA if we have the skills to bail out.   Without these skills it is riskier to make the commitment. Even if a decision maker does adopt a CoA, having to continually think about contingencies can be distracting from the task at hand. More experienced decision makers seem to be confident that they can recognize warning signs in time to react, so they don't have to make deliberate tests all along the way.

FUNCTIONS PROVIDED BY IDENTIFYING A COURSE OF ACTION

The most direct and dramatic reason to select a CoA is to resolve an incident.   For example, the decision to fire a missile at a threatening aircraft is intended to remove the threat. The decision to divert a commercial airliner to a closer airport, after a fuel leak is discovered, is intended to eliminate the risk of running out of fuel. In studying decision making in an AEGIS cruiser, Kaempf et al. (1992) found that these terminal decisions made up only part of the set of decisions about a CoA. There were at least two other classes of CoA decisions.

Many times, decision makers select a CoA to prepare for resolving the situation. They select actions for contingencies that may arise later, but these actions won't resolve the problem itself. One example of this was a Navy Commanding Officer (CO) trying to identify an inbound air track. The CO did not wish to mistakenly shoot down a friendly aircraft but also could not afford to put his own ship and crew at risk if the aircraft turned out to be hostile. Therefore, the CO

ordered the weapon system used to engage missiles to be prepared. This did nothing to improve the chances of identifying the unknown track but it did shorten the response time to take defensive action should it have become necessary.

Most common of all are decisions about managing resources. In a threatened AEGIS cruiser, should the crew activate anti-air missiles? The prudent commander will want to ready defenses while figuring out what an airplane has in mind. But some missiles can only be activated for a short time before their batteries run down. A commander who activates these missiles too soon will compromise their value, while waiting too long precludes their being ready when needed. So, this type of decision is about resources, not about whether to fire a missile at a possible threat.   Sometimes these types of decisions treat information as a resource. For example, Kaempf et al. (1992) report cases where the AEGIS commander and crew debated about whether to bring in a CAP to visually identify an unknown track. If the CAP could arrive in time, the ambiguity would be resolved.   But if it arrived too late, it would just delay the ship from defending itself, and the CAP might even get too close to the track to allow the AEGIS cruiser to fire its missiles, so it would be a hindrance, not a help. As you can see, much decision making goes into preparation and management, and not just ending the incident. We need to take this into account in designing systems because it is easy to focus on the dramatic part of the task, and spend less attention on the mundane CoA decisions that may actually be more critical and more easily supported.

STRATEGIES FOR SELECTING A COURSE OF ACTION: SINGULAR EVALUATION

There are two ways in which a decision maker can arrive at a CoA: by picking the first one that seems adequate, or by using comparative evaluation to contrast alternatives to find the best one. This section describes the first way of arriving at a CoA, by using singular evaluation.   This is a "satisficing" (as opposed to an "optimizing") approach, so named by Simon (1955) who observed that

people in business typically just needed a CoA that would do the job, and so needn't go to the effort of trying to figure out the best option. In fact, for most business decisions it would be impossible to determine which option is best, because the goals are ill defined.

Models of naturalistic decision making usually include a satisficing criterion. They see people as trying to quickly find a workable solution. Therefore, the strategies for selecting a course of action are typically singular strategies that look at the options one at a time, until an acceptable one is found. If an experienced decision maker identifies a workable option as the first one considered, the search ends right there, and no other options will even be considered. Why go to the trouble of generating more options, and then evaluating them, if you don't have to? This method works even if you don't generate a reasonable CoA as the first one considered, if you can quickly see its shortcomings and reject it. In general, even moderately skilled decision makers are able to generate a workable CoA as the first one they think of. Klein et al. (1992) studied medium-ability chess players, who were asked to think aloud while reviewing different board positions. The initial moves mentioned were of very high quality, far better than would be expected by chance. The experience of the players enabled them to truncate their search by homing in on the best moves right away.

We are used to thinking about option evaluation as evaluating strengths and weaknesses of alternatives, so singular evaluation is a little different. Singular evaluation strategies are ways of evaluating CoAs one at a time, until an adequate option is found. If that is the first option considered, the search can end right away. Skilled operators expect their experience will enable them to generate an adequate option as the first they consider, so they don't expect to be looking at many options.

A simple singular evaluation strategy is to check off if the CoA has the necessary properties. Kaempf et al. (1992) found this to be the most common form of singular evaluation, in their study of AEGIS anti-air warfare incidents.

A more complex strategy is to use analogical reasoning. An option can remind a person of a previous case, and the success and/or disappointment can be used to anticipate what might happen if the option were put into action.

Mental simulation is an effective strategy for evaluating a CoA, just as it is an important source of power for situation assessment. Klein (1989) reported that decision makers in a variety of domains would evaluate options by playing them out in their minds, looking for flaws, trying to find ways around the flaws, rejecting the option if it couldn't be repaired, and implementing the option if the mental simulation didn't turn up any fatal flaws. Mental simulation is a strategy for conducting a deep search of a few options, often just a single option, as opposed to the comparative strategies that are shallow assessments of a large set of options. Mental simulation allows the decision maker to consider an option in the context of the situation, and if difficulties are found, mental simulation helps the decision maker improve the option. While it may seem that a singular evaluation strategy is a sign of laziness—e.g., go with the first CoA you think of—a person using mental simulation can sometimes expend as much effort in deepening the search as would be required to generate and contrast different options. If experience enables you to generate a reasonable option as the first one considered, then you may be better off using mental simulation to evaluate that option than to try to think up other CoAs.

The RPD model, described above, covers all three of these singular evaluation strategies. The model was designed to explain how people can make decisions without contrasting different options, and each of these satisficing strategies accomplishes that function.

STRATEGIES FOR SELECTING A COURSE OF ACTION: COMPARATIVE EVALUATION

There are times when a decision maker will need to choose among different options. If your car breaks down, and the cost of repairs is greater than the value of the vehicle, you will have to buy a new car. This usually means you will visit different dealers to identify

alternatives and to pick the best. Or if you move to a new city, and your realtor tells you about several houses in your price range that are on the market, you will have to compare them. Community planners may have to choose the best site for a new airport from among several candidates. Fischhoff (1991) has recently studied the alternative courses of action available to women who need to defend against physical attack. Another such choice is the common one faced by high school graduates trying to pick a college to attend. Generally, it is when we don't have experience in a field that we have to select among options.

In choosing among options, the decision maker is performing a comparative evaluation (either feature-by-feature or option-by-option). The purpose of the evaluation is to contrast the alternatives and find the best one—to optimize. Following the lead of Svenson (1979), Zsambok, Beach, and Klein (1992) have identified 15 different strategies for choosing from multiple options. These are all strategies that break the CoA decision into components and features, and perform analyses of these features. They also reviewed more recent literature and identified a small set of strategies that people would be likely to use in naturalistic settings. They eliminated strategies from Svenson's list that required computation, or took excessive time, or required high data quality.

In performing an option analysis via features, the decision maker identifies the feature(s) of interest, lines up the different CoAs, determines the extent to which each CoA accomplishes the feature of interest, and uses this datum to compare the options. Zsambok et al. have identified some typical strategies: Elimination-by-Aspects, Conjunction, Disjunction, and single feature inferiority.

Elimination-by-Aspects[30] is a method of setting successive hurdles. The most important feature becomes the first hurdle. All options are evaluated on this feature, and any that fail to meet a criterion are deleted. The process continues until one option is left, and that is chosen. For instance, in hiring one of several qualified job candidates, the résumés that show fewer than two years of experience might be removed, and then the résumés showing no progression of responsibilities, and so on.

Conjunction is the use of several criteria that all must be met to select an option, e.g., buying a car that is inexpensive and also has an airbag. Disjunction is the use of several criteria, only one of which must be met, e.g., hiring an engineer who is trained either in structural engineering or software engineering. Single Feature Inferiority is a strategy for comparing a pair of competing options by rejecting the one that is worst on the most important feature, irrespective of its standing on other features of interest, e.g., rejecting a proposal that is much more expensive than the others, ignoring the quality of the approach. As you can see, these are all quick and dirty methods for arriving at a choice, not necessarily the best choice.

There are also two global strategies for selecting among options without using feature analysis. One involves mental simulation, and the other is a method of paired comparisons.

The Dutch psychologist Adriaan de Groot (1946/1965) was one of the first to show how mental simulation functioned. He collected think-aloud protocols from skilled chess players who were trying to find the best move in a complex board position. Rarely did the players contrast the strengths and weaknesses of different moves. Instead, they identified one move at a time, and used progressive deepening (which is de Groot's term for mental simulation) to imagine how that move would develop. They played the continuations out in their minds. If they found any flaws, they would search for ways to improve the line of play. If they could not find any way around the flaws, they would reject the move. The chess players carried out their mental simulations for several moves, forming a global reaction to each move as one that pleased them ("I'd like that position") or displeased them ("Take it away!"). They compared their global, emotional reactions of the moves to select the one with which they felt most satisfied.

Another strategy for contrasting options is the use of Successive Pairs.[31] Here, the decision maker selects two options from a larger set, compares them without necessarily decomposing them into features, selects the more appealing of the two, deletes the other, replaces it with another member of the pool, and repeats the process until all but one of the options have been rejected, and a favorite has emerged. This is an efficient strategy for canvassing all the

alternatives without ever having to hold more than two in memory at the same time. For example, a pilot who has to divert because of bad weather has to select an alternate airport. It is hard to compare all the possibilities, so the pilot might consider them two at a time. Airport A is better than B because A has longer runways, and B requires tighter maneuvers during the final approach. A is better than C because it offers more possibilities for passengers to make connections, even though the runways are equivalent. And the last option, D, is better than A because it has a depot where the mechanics can make some minor repairs, with everything else balancing out.

Finally, we should consider some decision strategies that are prescriptive. They are not included in Table 4 because they rarely would occur in naturalistic settings because they require a great deal of time and expertise, a high level of data quality, and usually some sort of decision support. These are analytical strategies for measuring with some precision the strengths and weaknesses of options on specific evaluation dimensions. These strategies are considered to be compensatory, because if an option has a severe weakness on one dimension, it will offset mild strengths on several others. One example is Multi-Attribute Utility Analysis[32], in which the decision maker rates each option on each pre-defined dimension, and also weights the evaluation dimensions. Another example is decision analysis.[33] These strategies seem to be good candidates for decision aids, but have not been well accepted because they require the types of inputs—probability estimates, anchored ratings—that are difficult to provide, and because they provide answers in terms of quantitative scores that users cannot easily interpret.

Boundary Conditions for Different Decision Strategies

When will a person use singular evaluation strategies, versus comparative strategies, and even the more analytical, compensatory strategies? It is important to realize that each type of strategy has its strengths and weaknesses. Singular strategies can leave the person fixated on a mediocre CoA, and missing a much better one. It is probably a mistake to use a singular, satisficing strategy such as RPD

to select a site for storing nuclear wastes. Comparative strategies can provide the illusion of rationality without capturing contextual features that are hard to analyze. It is probably a mistake to use a feature comparison evaluation strategy for deciding whether to abort a takeoff when an engine fails just as you are getting up to speed.

Table 5 shows some of the factors governing the use of strategies (adapted from Klein, 1989).

The singular, satisficing strategies, along with more overarching NDM strategies such as recognition-primed decisions, are most likely when time is restricted, goals are unclear, the decision maker has some task experience, and the conditions keep changing.

The comparative evaluation strategies, including prescriptive strategies, are more likely when a person needs to optimize, or at least needs to be seen as trying to optimize, when the person needs to document and justify the choice, when the data are abstract, when there are multiple stakeholders, and when the problem is combinatorial (e.g., interactions between different drugs). To use comparative evaluation strategies, the goals of the task should be well defined, the evaluation features must be pre-selected, and there should be guidelines for assigning weights and common metrics for making the judgments. The data quality needs to be high, and time pressure should be low.

In the hands of the best decision analysts, comparative evaluation methods, particularly the analytical, prescriptive strategies, are powerful tools for addressing complex, multi-faceted issues such as estimating risks of accidents in nuclear power plants, deciding where to locate airports, and deciding whether to pursue social policy programs. These are certainly real-world decisions even if they don't possess the NDM features listed at the beginning of this report, in Table 2.

There are times when comparative evaluation of options is not worth the effort. To a person making a choice among different options, the process can appear difficult and discouraging. Janis and Mann (1977) felt that most people avoid decision making if they can. Yet, for all the pain, the comparative evaluation decision itself may not be very important because in many cases the person is relying on only a small subset of relevant information. The deliberations are

Table 5. Boundary conditions for different courses of stategies.

	Singular	vs.	Comparative
Time Pressure	x		
Experience Level	x		
Dynamic Conditions	x		
Ill-defined Goals	x		
Justification			x
Conflict Resolution			x
Optimization			x
Computational Complexity			x

buried in the noise level. Consider a high school student selecting a college. The choice is clearly important, it will affect the student's choice of career, possibly the student's choice of a spouse, the place the student lives, and so forth. Still, the student can only consider a few variables, e.g., student-to-faculty ratio, types of electives offered, the pleasantness of the campus on the day it was visited. These variables are probably insignificant compared to the factors that will wind up governing choice of major, spouse, and location. In such a situation, the decision process may be a waste of time. If one option seems clearly better than the others, it is the one to adopt. But if no option stands out, then it doesn't matter which option you pick, because the variables you are considering are so trivial compared to the ones that matter. This interpretation is not comforting, and is intended as a caution against spending too much effort analyzing complex cases, when there aren't enough data and experience to make the exercise worthwhile.

These last three chapters discussed the situation assessment and course-of-action decisions that people make in naturalistic settings. We have examined the outputs of the decisions, the functions served by the decisions, and the specific strategies used to make diagnostic and CoA decisions. This chapter also covered singular versus comparative CoA strategies. By this point, you should have a good sense of how operators will approach a cognitive task. The next question is what can go wrong in a naturalistic setting, that might result in a poor decision.

FACTORS THAT CAN REDUCE THE QUALITY OF DECISIONS

Now we must examine the question of why good people make poor decisions. David Woods (1990) has argued that the field of human factors has shown designers how to eliminate slips and confusions in the use of control panels. According to Woods, the remaining challenge is to attack the cognitive errors so that designs can reduce the rate of inadequate decisions. These errors can arise for several reasons. Simple, mechanical equipment has become more complex with the introduction of computer technology, e.g., televisions and telephones. Additionally, computer technology allows operators to control more complex tasks, so the decision requirements are more severe. We are still learning how to satisfy these requirements.

There is ample evidence of poor decision making all around us. People select the wrong options, make errors, and show poor judgments despite training, careful design of equipment, alarms, and warnings. Michael Doherty (1993) has asked whether the NDM approach will ever be able to explain how people make bad decisions, since the field research and naturalistic observation methods are used to describe events, and not to evaluate the quality of the decisions.

On the other side, there are researchers and professionals warning us not to be too quick to attribute every system failure to a decision error.[34] Baruch Fischhoff[35] cites the example of an investigation into an airliner that crashed. The crash seemed to be caused by a mechanical problem. But after two weeks, the investigators were able to show that there was a way in which the pilot might have recovered

control. While they didn't list the cause as entirely due to human error, the report pointed to human error as one of the contributing factors. Fischhoff wondered how the investigators could have expected the pilot to find a solution during an emergency, when it took them two weeks of calm deliberation.

Jens Rasmussen[36] has been more emphatic on the difficulties of attributing errors. Once it is clear that something has gone wrong, the causal chains spread backwards to the operators, the people who trained the operators, the people who prepared the training program, the personnel selection staff, the human resources staff, the system being operated, the people who designed the system, the staff who maintained the system, the organization that set up incentives, safety standards, and so forth. Depending on where one sets the stopping rules, this chain of blame can continue indefinitely. From this perspective, it makes little sense to Rasmussen to conclude that an accident was caused by operator error. And it makes little sense to try to trace errors to the decision strategies used by the operators. Rasmussen also argues that we have become too fixated on eliminating errors. He claims that errors are inevitable as operators test the boundaries of a system. Such boundary testing builds expertise. Catastrophic failures can arise when operators are prevented from testing the boundaries, so that when unexpected breakdowns occur the operators lack the skill to respond. Rasmussen's suggestion is to put more energy into building robust systems that recover easily, and to build displays that show the operator the dynamics of the incident, rather than trying to build error-proof systems.

In this chapter, we will examine some of the ways NDM strategies can fail. But, siding with Fischhoff and with Rasmussen, we believe that poor outcomes do not mean there was poor decision making. It would be dramatic to announce some new reasoning failures that are tied to NDM strategies, but none have emerged thus far.

CAUSES OF ERROR IN NATURALISTIC DECISIONS

A review of decision errors that were committed during actual incidents by firefighters, commercial airline pilots, and intensive care

unit nurses[37] found that the most common cause was simply the lack of experience. The data had been collected as retrospective protocol analyses, in which people described nonroutine and challenging events. The database consisted of approximately 450 decision points studied. The decision makers acknowledged 25 cases where they had made the wrong choice, so the error identification was from individual admission. In case after case, the decision makers could say what they should have done, given the information available. But when asked why they hadn't made the right choice, the informants shook their heads and described a key relationship they hadn't understood, or a causal factor they hadn't appreciated. A total of 21 out of the 25 errors could be accounted for by lack of experience.

Example 6.1 Error attributed to lack of experience: The importance of roof construction in conducting firefighting operations

A relatively new fireground commander inspected a building that was on fire, and judged the fire to be controllable. Soon after, the roof became unstable, and the job became more difficult. The commander had not realized that the building used balloon construction which is economical and stable, but is vulnerable to damage to the supports. A relatively small fire compromised the stability of the entire structure. In retrospect, he said he should have noted that the building used balloon construction, and called in a second alarm right away, when there was a chance to save the supporting struts. After that event, the commander was careful to identify the type of building construction when called out to fires.

Example 6.2 Error attributed to lack of experience: Understanding the dynamics of a commercial airline cockpit

This incident comes from a study of commercial aircrews responding to malfunctions during a simulated flight.[38] There was a malfunction resulting in decreased oil pressure. The oil pressure dropped just above the point where the engine would

need to be shut down. The flight engineer decided not to shut down the engine, but to reduce power. The reduction in power cut back on the fuel flow. The reduction of fuel flow actually raised the temperature of the engine, because the flow of fuel helps to cool the engine. The flight engineer had been taught about this linkage, but had forgotten about it, and noticed that temperature was rising for an engine unit already having oil pressure problems. He decided to turn the engine off to prevent further damage. So, the engineer caused the extra symptom (high temperature) by his own action (reduced fuel flow) and then interpreted the symptom as a sign of an impending oil system failure. The problem seemed to be in the flight engineer's limited experience base that left him unable to make a connection to facts he had been given during training.

These two examples show poor decisions that were not caused by faulty logic or carelessness or information processing limits or biases, but simply by gaps in the decision maker's knowledge and experience. Because of these gaps, the decision makers did not see implications and relationships that would be clear to them in the future.

It should not surprise us that the most frequent problem was lack of experience. The NDM approach is about the use of experience. Strategies such as the RPD model are simple. The power of the strategy is not in the steps. It is in the fact that the strategy enables decision makers to use their experience. Operations researchers could easily simulate the RPD strategy, but that would accomplish little. The hard part would be to simulate the experience base of the skilled decision makers.

The review of instances of errors turned up a few other sources of error, but their frequencies were much lower and they overlapped the category of errors due to inexperience. In several cases, the person failed to acquire necessary information, either due to inexperience or to workload. The significance of the information usually became clear in retrospect. Another problem was being misled by an analogous case. Here, the decision makers missed some important differences,

essentially because they lacked the experience to understand why a prior case was a poor analogue. Of the 21 errors traced to lack of experience, seven involved misuse of analogies.

Table 6 shows potential weaknesses in NDM strategies linked to the three primary forms of inference used in NDM—feature matching, analogical reasoning, and mental simulation.

• For feature matching, the possible mistakes are using the wrong features, or using inappropriate weights for features. These do not appear to be reasoning errors (e.g., logical fallacies), but limitations in experience.

• For analogical reasoning, the possible mistakes are selecting the wrong analogues, or making the wrong adjustments in adapting the analogues to fit the current case. Again, these are more related to inexperience than to reasoning errors.

• For mental simulation, there are several possible mistakes.[39] One problem is to build an inadequate simulation, due to lack of experience or lack of time. Another of these problems is to hold on to a mental simulation even when it is contradicted by experience, by explaining away the inconsistent data.[40] Marvin Cohen (1991) has argued that it makes sense to explain away some apparent inconsistencies, but when the weight of inconsistencies starts adding up, and the amount of explaining away becomes too great, the decision maker must be able to abandon the explanation. However, if the decision maker is not keeping track of how many ways the simulation has been patched up, he or she might fail to notice that the simulation has been running into trouble.

In reviewing these weaknesses in NDM strategies, all of them seem obvious and straightforward. There are no critical breakdowns in reasoning, no obvious thinking disabilities. The most general problem is that decision makers sometimes lack the experience to handle nonroutine incidents, but this finding doesn't give designers guidance for helping decision makers in the same way designers have learned to reduce motor slips by giving switches different shapes. Suggestions have been made to provide supplemental information on the display, to remind the operator of important facts or cue important relationships. This type of aid can do more harm than good. It clutters up the display, and it irritates experienced operators who don't

Table 6.   Limitations in the use of naturalistic reasoning strategies.

Feature matching   Using the wrong features

                              Weighting the features incorrectly

Analogical reasoning   Selecting the wrong analogues

                              Adjusting the analogues incorrectly

Mental simulation   Building an inadequate mental simulation

                              Failing to reject an inaccurate mental
                              simulation

                              De Minimus

                              Fixation

need the extra information. A potential fix is then to try to build an adaptive interface that would sense the operator's skill level and mission need, and present the information only when appropriate. Currently, adaptive interfaces have not been successful. Expert systems technology struggles with simple domains, and is not capable of reliably reading the mind of the operator.

As designers learn to identify critical decisions, and learn to anticipate how a decision maker will be drawing on information and comparing different trends, the system design should improve by reducing memory strain and workload, and error rates should go down. The improvements should come from using decision requirements to support cognitive systems engineering.

Heuristics and Biases

In looking for sources of decision errors, one important line of research is the attempt by Kahneman, Slovic, and Tversky[41] to identify the reasoning heuristics people use. In addition to illuminating the process of thinking, this inquiry can also show us where the heuristics commonly fail. Look at it this way; an algorithm is a procedure designed to churn out a correct answer to a problem. Algorithms may be slow, tedious, and inefficient. Heuristics are short-cuts for arriving at answers without going through all the steps of an algorithm. But they don't guarantee a correct answer. They work most of the time, which is why people use them. Sometimes, the heuristics fail and lead the decision maker in the wrong direction. So they act as biases. The efficiency of a heuristic for restricting search is also a bias that prevents the decision maker from noticing certain types of cues. According to the logic of this approach, if we can find out where heuristics don't work, we can figure out ways to design systems to alert the operators that their biases are getting them into trouble.

Andrew Sage (1981) has carefully reviewed the decision heuristics/biases studied and identified approximately 25 different types. A few common examples are availability (classifying a situation according to a category that is readily remembered),

representativeness (stereotyping a situation), anchoring and adjustment (starting a diagnosis by classifying the situation in a certain way, and then making small adjustments to the estimate with each new piece of information, rather than re-classifying the situation), and confirmation (seeking information that confirms a hypothesis rather than looking for information that might reject it). Each of these heuristics is valuable, to guide memory search and classification and appraisal. And each heuristic can result in the wrong answer, under certain conditions.

The heuristics/biases research seems to have clear implications for helping people make better decisions. Russo and Shoemaker (1989) have written a book *Decision Traps*, describing how common heuristics/biases can lead business executives astray. Russo and Shoemaker also present antidotes for the heuristics/biases.[42] Zachary, Zaklad, Hicinbotham, Ryder, Purcell, and Wherry (1991) have discussed the importance of designing systems and interfaces that can protect the operators from decision biases.

Unfortunately, the heuristics/biases approach has not been very useful when it comes to practical applications. Lopes[43] has explained that the experiments used to demonstrate a heuristic had to be so designed that if subjects were using the heuristic they would get the wrong answer to a formal problem. This procedure made the demonstration more convincing; it showed that subjects used the heuristic even when it misled them. But because all these studies of heuristics kept showing subjects making errors, they conveyed a general impression of people as flawed decision makers. Many researchers saw the opportunity to improve decision quality by helping people avoid the biases.

What researchers didn't ask was how likely these heuristics/biases are to result in poor outcomes. The answer seems to be that the biases aren't worth worrying about. Christensen-Szalanski (1986) found that in medical diagnosis of pneumonia, there was a bias in estimation but it had little impact. A given practitioner might miss one case in a year, and that miss would be corrected when the patient failed to improve. Shanteau (1989) has concluded that experienced accountants don't show the common biases, and that as people gain experience in a field, the likelihood of decision biases diminishes, so these biases do not seem to be built in to the way people make

decisions. Fraser, Smith, and Smith (1990) have found little evidence for the various biases in naturalistic settings. In studying the confirmation bias, Klayman and Ha (1987) found that there were situations in which confirmation seeking was a reasonable strategy. Gigerenzer, Hell, and Blank (1988) showed that certain evidence for biases might be artifacts of the experimental design. For example, Tversky and Kahneman (1983) asked subjects to estimate probabilities, and rigged the experiment so that if subjects used representativeness, they would arrive at the wrong answer. Gigerenzer et al. speculated that the subjects suspected the experiment was rigged, so they didn't bother estimating probabilities. When Gigerenzer went to the trouble of showing pieces of paper being drawn randomly, the subjects used probabilities, and the representativeness bias diminished. Gigerenzer and Murray[44] and Cohen[45] have presented theoretical analyses criticizing the decision biases concept.

In summary, the decision biases approach does not seem to be useful for design engineers as a source of guidance in reducing errors. Why bring it up in this report? Because the body of work on heuristics and biases is very well known, and its omission would have bothered some readers. And because the paradigm, having generated so much research, might yet demonstrate applied implications.

Stress and Decision Making

One of the features of naturalistic settings is the presence of acute stressors[46] that are sudden, unexpected, and of short duration. Acute stressors include:
- time pressure
- ambiguity
- noise
- threat
- impending failure
- public scrutiny of performance
- high workload

These conditions make it difficult to carry out many analytical strategies, and may also disrupt decision making in general.[47]

How can acute stressors affect decision making and cognition? Table 7 presents some possible mechanisms, seven possible ways in which different acute stressors can have an impact on decision processes, either by changing the internal cognitive resources, or changing the criteria for using different strategies. Some of these mechanisms are specific to a single stressor, but most are possible reactions to a wide variety of acute stressors. Noise can interfere with inner speech, making it hard for people to think to themselves. Threat can set up a secondary task that distracts people from the primary task. For example, physiological symptoms such as shortness of breath and trembling can require a secondary task to manage the symptoms. Increased self-monitoring, to see how you are holding up under the threat, also is a secondary, distracting task. Further, people under stress may be less able to use working memory, if only because of the secondary tasks to be handled. Narrowed attention can affect decision making. A number of studies have shown that people under stress examine fewer cues, perhaps because there is not sufficient time, or because the stressor interferes with attention.[48] Hammond (1990) has suggested that the narrowed attention may be an effective means of focusing attention on the most important cues.

There are additional reactions to acute stress, and some of them may be adaptive under the circumstances. These are the last four entries in Table 7.

Finally, the nature of the task may create pressures to increase speed at the cost of accuracy, or vice versa, so criteria will shift.

Under stress, people may rely on simpler strategies for selecting CoAs. Payne, Bettman, and Johnson (1988) found that, under time pressure, it was more effective to shift to simple, noncompensatory strategies than to persist in using analytical compensatory strategies. And, of course, the singular evaluation strategies described earlier can be used with much less time and effort than the comparative evaluation strategies.

Preferences can shift, as people become more conservative (Edland, 1989). This shift towards conservatism may be reasonable, given the reduced capability for adjusting and improvising. People under stress may show greater rigidity, and fixate on their initial situation assessment even in the face of contrary evidence.[49] Even

Table 7. Some ways in which acute stressors can affect decision making.

Interference with inner speech

Secondary distracting task

   Increased self-monitoring

   Reduced efficiency of working memory

Narrowed attention

Speed/accuracy tradeoff

Use of simpler decision strategies

Conservatism

Fixation

this tendency may be adaptive, because a person may become confused from entertaining different hypotheses when placed in a distracting environment.

The list of stress effects presented in Table 7 constitutes one of the motivations for research into NDM. Even for tasks where quantitative, analytical decision strategies are appropriate, these stress effects will make it difficult if not impossible to carry out such strategies. It seems reasonable for people to use the simpler, singular evaluation strategies, and to use experience in sizing up situations to avoid option comparison altogether.

Moreover, the simpler, naturalistic strategies such as recognition-primed decisions, do not necessarily result in poor decisions. In many studies, high levels of acute stress did not disrupt decision quality and sometimes even increased decision quality.[50]

You can use the effects listed in Table 7 as a checklist to help imagine how the tasks and sub-tasks will feel under operational conditions. An interface that makes sense back in the office may not work so well in the field. Interfaces that do not reflect these constraints can make it difficult for operators to perform their jobs, usually at the most critical moments.

Example 6.3 Ignoring time pressure: JSTARS

Returning to the earlier discussion of the self-defense suite of the JSTARS aircraft, you will recall that the interface was built using a menu structure that was generally two to five levels deep, and at one point, 19 levels deep. This was unacceptably cumbersome, because one of the most critical decisions was whether to continue the mission or break it off because of a threat. This decision was going to be made under extreme time pressure and required an interface design that was more efficient than a menu structure, especially one with successive levels.

Had the decision requirements been carefully mapped out in advance for the self-defense suite, it might have been possible to anticipate that the interface design needed to be improved. In fact, it should have been possible, during the

conceptual phase, to anticipate the type of strategy that the operator would be using. The CoA decision was going to be less important than the situation assessment decision. That is because the CoAs were fairly limited, and followed directly from the interpretation of level of vulnerability. The situation assessment decision would be to determine when the threat became too great to ignore. Because the self-defense suite operator was working as a team with the pilot and the mission commander, the self-defense suite operator was not necessarily going to make the situation assessment decision, but was going to keep the others informed as the situation deteriorated.

The difficulties in designing the JSTARS self-defense suite cannot be attributed to the system designer alone. The DoD personnel played their own role. They identified AWACS as the analog aircraft, because AWACS also flies a surveillance as well as a command-and-control mission. However, AWACS is concerned with the air picture, whereas JSTARS is designed to convey the ground picture. AWACS flies well behind the battle lines, whereas JSTARS needs to fly closer to the battle lines. AWACS has dedicated CAP, whereas JSTARS relies on AWACS to provide CAP support. AWACS was a useful analogue in some ways, but was a misleading analogue in other ways for the self-defense suite of JSTARS. The fact that AWACS seemed safe against attack was used as evidence that JSTARS would also be safe, and didn't have severe self-defense needs. Some people familiar with the JSTARS project did worry that the airplane was such an easy target it would be continually breaking off its orbit, and therefore would be unable to perform its mission, but this was a minority opinion.

During interviews, DoD design personnel, when asked about the high workload facing the self-defense suite operator, argued that there would be plenty of help. On board the AWACS, there are several Air Surveillance Officers and Technicians who have radar scopes scanning different sectors of the sky. It was argued that on JSTARS, all these

surveillance personnel would be able to provide useful information, just as they did aboard AWACS. What was missed was that in AWACS, the Air Surveillance personnel are all looking at the sky, trying to pinpoint the locations of enemy and friendly aircraft, whereas in JSTARS, all the surveillance radars would be directed at the ground, and would be oblivious to the air picture. This shows how captivating and misleading the AWACS was as an analogue to the JSTARS self-defense suite.

It is critical for those in the early phases of the acquisition cycle to closely examine the differences between those systems they have identified as representative "predecessor" systems and the system they are proposing to be developed, to determine the likely implications of the differences between them.

The JSTARS example shows how time pressure needed to be considered as a stressor in designing the subsystem and interface for self-defense.

The next example has been widely cited as showing the effect of stress on decision making, but the problem may have had more to do with the interface than the decision processes.

The Vincennes shoot-down is an interesting case to examine, because with hindsight we can say that the wrong decision was made; it contains elements of stress (personal and professional risk, time pressure, noise, ambiguity, high workload, and public scrutiny of performance), and it has been cited as a demonstration of decision bias. In addition, the interface design has also been blamed for the difficulties the crew members had.

Example 6.4   An uninformative interface: The Vincennes shoot-down

On 3 July 1988, the USS Vincennes, an AEGIS cruiser, mistakenly shot down an Iranian Airbus, killing all of the 290 passengers. The Navy issued the Fogarty report, the official account of the incident. The Fogarty report identified stress

as a contributing factor to the mistaken decision to fire missiles at the Airbus, and also pointed out that the AEGIS displays contributed to the problem. Several behavioral scientists,[51] in testifying before a House subcommittee, attributed the mistake to decision biases.

The Vincennes incident has been the focus of many articles, and at least one book (Rogers and Rogers, 1992). Four years after the episode, interest was still sufficiently high to justify a Newsweek (1992) cover story. The incident also generated a major research program by the Navy: Tactical Decision Making Under Stress (TADMUS) to learn how to avoid future decision errors through better interface design, decision support systems, and training.

The background of the incident is that the Vincennes was part of a U.S. effort to keep the Persian Gulf safe for commercial shipping during the Iran-Iraq war. It had been harassed by Iranian F-4s several months earlier while carrying out a routine escort mission (see Example 3.1). Iranian F-14s had been shifted down to the Bandar Abbas airport (which was used for military and commercial purposes) two weeks earlier. In addition, there had been a recent incident in which an Iranian F-14 had flown towards a U.S. cruiser, had been warned, and had broken off and another incident where a U.S. Navy ship had to fire missiles at an Iranian F-4 that kept approaching it despite radio warnings. There were also instances in which Iranian F-4s had flown just underneath commercial airliners, for concealment. And intelligence reports indicated an Iranian action during the 4 July weekend.

The details of the incident are:

* *On the morning of 3 July, 13 Iranian gunboats had surrounded and attacked the USS Elmer Montgomery.*

* *The USS Vincennes had gone to the rescue of the Montgomery.*

* *The Vincennes had sent a helicopter over to the Montgomery for surveillance and support, and the gunboats had fired on the helicopter.*

- *When the Vincennes came into the area, it was also attacked by the gunboats.*
- *During the battle with the gunboats, the Iranian Airbus took off from Bandar Abbas airport.*
- *The Vincennes identified the aircraft as Unknown, Presumed Enemy because it took off from an Iranian airport used by military as well as commercial aircraft.*
- *The Vincennes received unconfirmed information that the aircraft was possibly an F-14.*
- *The Vincennes, fearful that an Iranian fighter was entering into the battle, attempted to warn off the aircraft using commercial and military radio frequencies, to no effect.*
- *The Vincennes determined that the aircraft was starting to descend towards it.*
- *The commander of the Vincennes ordered that the aircraft be engaged by missiles.*

One interpretation[52] is that the crew of the Combat Information Center of the Vincennes showed expectancy bias in judging the altitude change. In fact, the Airbus never descended towards the Vincennes. It continued to climb during the entire episode. The AEGIS data files confirm that the Airbus never descended. The crew of a nearby U.S. Navy ship saw only a steady ascent; they were not deceived. Yet several of the crew members on the Vincennes announced that the Airbus was descending, and no one contradicted this assessment even though the actual data were available at each work station. The decision bias hypothesis is that everything that preceded the incident had prepared the Vincennes' crew to be attacked. They were expecting an attack from Iran. And so the crew members distorted the altitude data to conform with expectations. They saw what would have happened if a fighter actually was diving towards their ship. According to this hypothesis, decision makers are prone to such distortion, particularly under stress, so it comes as no surprise. Here, at last, is a smoking gun showing where biases can lead.

The expectancy bias interpretation is dramatic but is not supported by the data. The altitude data in the Vincennes are presented as a four-digit alphanumeric readout on a side display, called the Character Read Out (CRO), not the primary graphic display boards. Worse than that, the CRO is filled with alphanumeric data for track number, speed, heading, and so forth. Worse yet, the CRO doesn't display trends, and neither does the primary display. The crew members have to study the digital read-out for several seconds, allowing for air pockets and system cycle time, before they can perceive trends. By some estimates, it takes 5-10 seconds of staring at the CRO to see a trend. Noise (coming in on separate channels for each ear, plus speakers in the Combat Information Center itself) adds further distraction to this task of trying to remember previous digital readings to infer trends.

There are other possibilities as well. The crew member who announced a decreasing altitude may have confused altitude with the range data, which were decreasing. Another explanation is that the error was due to a change in the track number. The Vincennes gave the unknown aircraft the track number 4474, but the system changed that to 4131, which had been used by another Navy ship. There is speculation that the crew members, by mistake, punched in the original track number to get information. Unfortunately, the track number 4131 had been reassigned to another airplane in the general vicinity, an airplane that actually was descending at the time. If this is true, then the crew members on the Vincennes read the trend accurately, but were looking at data for another airplane.[53]

In short, the hypothesis that expectancy bias was to blame is highly speculative. There are more obvious problems, particularly the design of the interface. Under these circumstances, the Vincennes needed an interface that showed altitude trends. It might have been helpful if the commander and crew could have seen changes in trends, to note if/when the unknown aircraft suddenly altered its climb.

The designers of the interface had envisioned a different mission. The AEGIS cruisers were developed to help the U.S. Navy counter large-scale attacks on the open seas. They are built to track and engage a large number of targets simultaneously, under conditions of declared war in which any track that was not identified as a friend was considered to be an adversary. AEGIS cruisers are not built to patrol coastal waters, or to figure out the identity and intentions of unknown tracks. So the AEGIS cruiser was not a perfect choice to be in the Gulf, on a mission of deconfliction. Yet it was the best ship in the Navy to defend against the silkworm missiles that Iran had recently obtained from China. So there was a tradeoff of defense capability versus sensor display capability. Only with the benefit of hindsight does the tradeoff seem wrong, and even then we haven't factored in the incidents and attacks that the AEGIS cruisers may have prevented.

The decision strategy the Vincennes' commander was using was probably constructing a mental simulation to make a diagnosis.[54] The cues were all consistent with a hostile aircraft: there was the Identify Friend or Foe indication of an F-14 (which probably came from an Iranian military airplane at Bandar Abbas while the Airbus was climbing after takeoff), the failure to heed radio warnings, the failure to fly straight down the center line of the air corridor (as virtually all airliners did), the timing of the takeoff during the battle with gunboats and not in accordance with flight schedules (the Airbus took off 27 minutes late), and finally, the descent. Everything fit.

In contrast, the story that the track was a commercial airliner ran into difficulties. Why would it ignore the radio warnings? Why would it deviate from the center line? Why would the Air Traffic Controllers send it directly into a surface battle?   Why would it descend when it reached the Vincennes? It is hard to build a story that explains these, and so the hypothesis that the track was a commercial airliner was rejected, without much difficulty. Some of these questions still aren't answered.

The Vincennes incident does not seem the result of expectancy bias. Even if the altitude trend had been accurately perceived, it is likely that the Airbus would have been shot down, because it continued to close on the Vincennes, ignoring radio warnings, and because a hostile fighter might have continued its climb until it was inside the engagement zone where the Vincennes defenses could operate.

The Vincennes incident may not even be attributable to time pressure. Extra time would have helped the crew gather more information, but given the information available at each point along the way, the commander and crew may have elected to do the same things even if they had more time. So, the real need was not for more time, but for more information or, at the very least, information that was better displayed to let the crew perceive what the Airbus was doing.

At this point in the report, we have surveyed the Naturalistic Decision Making approach, noting the characteristics of environments that are of interest, sketching out the nature of the decision strategies that would be used in naturalistic settings, for both diagnostic decisions and CoA decisions, and examining the factors such as stress that might affect decision making. The Vincennes incident describes an operational environment in which many of these factors came together.

What are the design implications of NDM? If you are able to consider the diagnostic and CoA decisions that operators of a system will make, how will that change your approach towards development? That is the topic of Chapter 7.

APPLICATION OF NATURALISTIC DECISION MAKING TO DESIGN TASKS

To illustrate the concepts of NDM, this chapter presents some ideas about how it might be applied for different tasks: diagnosing a situation and controlling a reaction.

DIAGNOSING A SITUATION

Diagnosis is important in many domains, such as:
- Medical diagnosis
- Troubleshooting a piece of equipment
- Accident investigation

Let us choose one domain to examine—the control room of a petrochemical plant, facing the task of figuring out what is happening in one of the processing tanks. This decision is about a course of action (whether to shut down the process), but it primarily is a diagnostic decision to understand why the problem has arisen. Is it a faulty gauge? Is it a minor impurity in the mix that will pass quickly? Is it an error in the feed, so that the mixture rate is too volatile? Once the diagnosis is made, the CoA will become clear.

This diagnosis decision fits within the parameters of NDM, as presented in Table 2.

- Time pressure is great, because the decision may have to be made within a few minutes of discovering the problem.

• The goals are not entirely well defined. Of course, the overall goal is to understand what is causing the problem and to take corrective action. This is a loose but sufficient definition. However, the sub-goals are not well defined because they are not yet known. It is during situation assessment that some of these will become clear. Experts would disagree about what is the correct course of action. Factors involved are the severity of the problem, the skill of the operator, and even the quality of the equipment, because that affects the amount of pressure the pipes and valves can withstand, which in turn can be a function of the competence of the maintenance crew in spotting defects.

• Goals shift during the incident. First is to minimize the problem, then to safely and quickly shut down a tank, and then to rapidly shut down the plant, and even to evacuate the area.

• Ambiguity is present. The readings from different gauges aren't always reliable, and some gauges may cease to work. Or else temperature gauges may be unreliable in certain ranges. Because temperature gauges are located along the walls of the tank, the operator won't know what is happening in the center. Or, the operator may not know where different chemicals have moved during the cycle, so that a temperature that is acceptable for one chemical might be alarming if a different chemical has moved to that area.

• Experience is critical, and the cost of training and retaining experienced operators and supervisors can be a fraction of the cost of an accident, or of repeatedly shutting down the plant.

• This is a team decision, involving the supervisor, the operator in charge of the tank, and the workers out in the field who have been directed to open and close different valves.

• Organizational precedents are important, showing the conditions under which shut-downs were rewarded and punished.

• Stress is obviously present and there are other, higher-level goals to consider (e.g., preserving safety records, loss of credibility from a false alarm).

• The procedures for diagnosing malfunctions are poorly specified.

• The risks are high, because shutting down the reaction is expensive in terms of lost production, and because the cooling down

of the asphalt lets it harden so that it has to be chipped out of the pipes and valves. Failing to shut the reaction down can result in a fire or explosion.

According to the NDM framework, there are several strategies for inferring a diagnosis. One is feature matching, which is certainly useful for the alarms, and even for developing a rule-based system to suggest diagnoses. However, the diagnosis problem isn't well suited to intelligent technology. The time course is so rapid that the operators would not be able to enter data into the expert system and be able to recover if the system could not provide a diagnosis. Moreover, expertise is needed to accurately enter the data, and operators with this level of expertise would probably be better off using their own judgment. The operators might be able to make better use of a RPD support system, such as Noble's, which would collect and interpret data to show which hypothesis is more in line with the actual events.

A case-based reasoning support system might be helpful, for calling up similar incidents so that the operator could map the current trends against those seen in the past.

In trying to diagnose the problem, operators usually are attempting to construct a mental simulation of what must be going on inside the tanks. They are trying to build a story, using the observed data, along with other knowledge (e.g., "the last batch from that refinery was deficient in certain ways," or "that valve has been known to stick"). The diagnosis is the story that makes the most sense to them. To help them build this story, an interface might do the following:

• make it easy to refer to the history of the event, so they can review the time course, and trace back if one event occurred before another.

• show when different control actions were made, because a frequent source of confusion is to misinterpret control actions as additional symptoms of the accident.

• make it easy to directly compare different trends. Too many interfaces force the operators to flip back and forth between different screens to make the necessary comparisons.

• help operators construct different stories, so they don't get locked into one because of memory limitations.

• help the operators construct and keep track of stories involving more than one malfunction, because multiple failures create the most confusing sets of symptoms.

• allow operators to notice when they have been making too many patches to repair a story. According to Cohen (1991), it is reasonable to fix a story to take some contradictory evidence into account. (E.g., If my story is true, why would that pressure be so low? Maybe the gauge failed.). But when you need to make too many repairs to the story, it is time to find a new explanation. Cohen argues that, under stress, people may lose track of how many deviations they have explained away. This has also been called the garden path problem,[55] because you keep going down the same path, fooling yourself that it is the right one. Displays that help an operator notice all the discrepancies may signal that it is time to look for a new path, a new story.

Note that these suggestions are not to push beyond the state of the art, to develop new concepts in Artificial Intelligence. Rather, the intent is to find ways to display the time course of the event so that the operators themselves can quickly construct and evaluate stories in arriving at a diagnosis.

The foregoing discussion illustrates ways you can support diagnostic decision making. These methods have relevance to other diagnostic tasks. For instance, Politser (1989) has studied the decision requirements for specific types of medical diagnosis. The standard display formats for test data were making it difficult for physicians to see certain relationships, but it was easy to design improved display formats. Turning to accident recovery, Rasmussen (1985) has examined the implications of NDM for nuclear power plants, showing how these large-scale systems have relied on a "defense in depth," to do everything possible to avoid accidents, and consequently have made the job of diagnosis even more difficult because the operator will need to troubleshoot the accident prevention system along with the plant itself, to untangle the pattern of alarms.

CONTROLLING A SITUATION

Control is important in many domains such as:
- Medical treatment
- Accident recovery
- Air Traffic Control

Let us continue with the example of the control room of the petrochemical plant. Once the diagnosis was made, the operator may have decided to shut down the plant. Now it is ready to start up again. The more quickly it can get back on line, the better. In a large plant, the operator may need several days to get certain processes going. The difference between a rapid startup and a slow one can amount to hundreds of thousands of dollars. But the more quickly the operator starts up the process, the greater the chance of having another problem.

During the startup, the operator will be faced with many CoA decisions, such as when to start certain processes, and which feeds to use. The operator may even have to decide what product to make, because the impurity of a given chemical may mean it is unsuited for certain purposes. How much heat and pressure the operator feels is safe can determine what may be done.

Again, these decisions fit within the parameters of NDM in Table 2.

- The risks are high, because a slow startup is expensive, and an aborted startup is even more expensive.
- Time pressure is great, because the process has to be controlled on line.
- The goals are fairly well defined, but there can be disagreements over the correct course of action just as with the diagnostic decision. The judgment of what is a correct CoA depends on the severity of the problem, the operator's skill, the quality of the equipment, the hazards posed by an accident, and so on.
- Goals may shift during the startup, from getting the process going, to trying to make a high quality product, to admitting that the current batch will be discarded.
- Ambiguity is present, in the form of instrument reliability and the absence of data at critical points.

- Experience is critical, because operators may have few chances to control a startup process.
- This is a team decision, involving the supervisor, the operator in charge of the tank, and the workers out in the field who have been directed to open and close different valves.
- The organizational culture defines which decisions the operator can make, and which must be deferred to a higher authority.
- Stress is present.
- The startup procedures may seem straightforward, but rely on complex relationships between cues so the judgment of when to move on to the next step is quite challenging.

According to NDM research, in making CoA decisions operators will not be contrasting different options. Instead, they will be needing help in developing and evaluating options. For simple cases, they will probably use feature matching to determine whether a CoA satisfies certain requirements, but simple cases should not require support.

We can think of feature matching in another way—as defining the critical features that need to be displayed. For example, in several cases discussed in this report, operators seemed to need to track changes in the rate of change of certain parameters. In the aviation incident where the flight engineer was monitoring a fuel loss, a critical feature was not just the fuel flow, but changes in the fuel flow for the wing tank that was leaking. In the Vincennes shoot-down case, one useful piece of information might have been not just altitude trends, but indications of changes in the altitude trends. These are second-derivative cues, changes in the rate of change, accelerations as well as velocities. By identifying the critical features and relationships, and making sure they are displayed, we should be able to give the operator a much greater level of control.

Analogical reasoning can be important for generating and evaluating CoAs. In some domains, such as system design,[56] people make extensive use of previous instances to come up with a planned course of action. In domains such as manufacturing, people make extensive use of analogues to envision how a product will be made, and to estimate time and costs. For process control, a case-based reasoning system might help an operator review previous startups, generating expectancies about trends and relationships that can be

matched against the current pattern, to give the operator a basis for noticing discrepancies as early as possible. The system can display templates that are reconfigured for the conditions under which the startup is occurring.

Mental simulation is needed to anticipate the consequences of a CoA. During startup, operators need to imagine what might happen if they maintain a certain level of temperature for more than a few minutes. Sometimes they need to imagine how a sequence of steps will take place, to notice inconsistencies or to detect ways in which their plan might be vulnerable. To support mental simulation, interfaces might include some forms of predictor displays. For planning an approach to start up the system, the operators might find it useful to work from a simulation format that lets them see the expected sequence of events, and prepare for contingencies.

This chapter was intended to illustrate some of the ways that NDM might enter into the design process. The next question is how to collect the data and perform the analyses that let you define decision requirements to design a system, interface, or decision support. The first seven chapters covered what NDM is, what the naturalistic decision strategies are, and how they are relevant to design. For the following two chapters, the report takes on a more practical tone. Chapter 8 describes methods of Cognitive Task Analysis, to enable you to gather information about how an operator performs a cognitive task, so you can define decision requirements. Chapter 9 follows up with ideas about how to transform decision requirements into design concepts.

COGNITIVE TASK ANALYSIS

Frequently this report has stated that it is important to understand the operator's decision-making strategies and inferences. Chapter 9 presents a Cognitive Systems Engineering process that can make use of decision requirements. But first we must consider what it means to understand a person's strategies and inferences. Cognitive Task Analysis is the attempt to get inside a person's head, to learn what he or she is thinking about in performing a task. Cognitive Task Analysis lets you define the decision requirements for a system. This chapter describes some of the methods used in Cognitive Task Analysis.

Cognitive Task Analysis tries to describe the way a person experiences a task and actually performs it. The objective is to determine:

- the key decisions
- the cues that enter into the decision
- distinctions between cues that appear similar
- the types of inferences involved
- strategies for making these inferences
- contextual factors that affect the inferences and decisions
- categories used to classify situations
- sources of confusion
- types of knowledge gained through experience

Conventional task analyses are carried out to list the steps needed to complete a task. A typical task analysis is finished when a complete set of steps has been identified, along with the criteria for

initiating a step and for judging that the step has been accomplished so that the next step can begin. Therefore, the difference is that task analysis is directed at the objective performance, whereas Cognitive Task Analysis is directed at the psychological processes underlying the performance. Task analysis tries to describe the objective signs of when to begin and complete a task or sub-task. Cognitive Task Analysis is also interested in the subtle cues that may depend on context and experience.

There are many technicians and professionals who claim they are just doing conventional task analysis, but have become skilled at asking about decisions, inferences, and tricks of the trade. They are doing a Cognitive Task Analysis without using that term. What you call your data-gathering method is less important than how you carry it out.

The reason for using the new term is that task analysis doesn't require the data-gatherer to probe about cognitive processes. And many people just carry out task analyses without going deeper than they have to. They may not know that they are staying at a shallow level, especially if they haven't seen a Cognitive Task Analysis. Another advantage of Cognitive Task Analyses is expense. Task analyses are straightforward and mechanical, but they are not easy. They require much work to collect and organize all the data. In contrast, Cognitive Task Analyses do not exhaustively review all facets of the task. Instead, they cut to the chase; they go after the critical cues and decisions, the things that distinguish experts and novices, or spell the difference between success and failure in using a system.

Basically, a task analysis is an algorithm. It is a sequence of steps that, if carried out as described, will lead to a given outcome. Too often, the steps are artificial. They may include activities that no one performs, that were added for logical consistency. They may miss short-cuts that are often necessary, but are hard to describe. Therefore, task analyses, as algorithms, may bear little relationship to what people are really doing. Designers can get into trouble using task analyses, because their systems may force operators to carry out sub-tasks that are irrelevant or obsolete. Task analyses seem to become increasingly inaccurate as tasks move from the procedural level to the cognitive level. The methods for conducting task analyses

were not designed to capture judgments and assessments, so the approach doesn't generate decision requirements.

Example 8.1   The difference between a task analysis and a cognitive task analysis: Following a recipe for a meal

A recipe is a task analysis. It lists the ingredients you need, and the steps you must follow in preparing a dish. To understand what is missing, imagine that you need to cook a dish for an organization you belong to that is holding a pot luck dinner. Imagine further that you have depended on your spouse to do the cooking, but he or she is unavailable at this critical time, and that it is important that you do an impressive job. And imagine that a friend gives you a recipe for an outstanding dish. There is just enough time to get to the supermarket, purchase the ingredients, bring them home, cook the dish, and get to the pot luck. You are tempted.

But think about what the recipe doesn't tell you. It doesn't tell you how to improvise in case one of the ingredients isn't available, so you won't be able to make a substitution. It doesn't tell you how to improvise in case one of the ingredients is atypical, e.g., drier than usual, or less ripe, so you won't know when to add more water, or cook for a longer time. It doesn't tell you what a component looks like when it is finished cooking, so you won't be able to tell when to take it off the stove. It doesn't tell you how to assemble the utensils so you may get caught short of pots and have to quickly wash one out while something may be burning. It doesn't tell you how to adjust in case your cooking utensils don't exactly match the ones listed in the recipe.

All of these types of knowledge are won by experience and perceptual learning. They are extremely hard to describe, e.g., how to explain to someone when a dish is done cooking. Philosophers call them "tacit knowledge,"[57] because it may be impossible to articulate them, especially if they depend on context. These are the types of

knowledge that make up expertise, and they are the target of Cognitive Task Analysis. They are commonly omitted in task analyses. Without knowing these essential aspects of performing a task, a promising recipe may become a recipe for disaster.

Cognitive Task Analyses are not intended to replace conventional task analyses. It is very helpful to study data-flow diagrams and task analyses before performing a Cognitive Task Analysis. When these conventional analyses are not available, because of time or expense, there are short cuts to learning what the task involves. Cognitive Task Analysis methods such as Concept Maps (discussed below) provide background information. Another way to prepare for a Cognitive Task Analysis is to observe people performing the task, or even trying it yourself.

Frequently, the operators or subject matter experts will be skeptical that an outsider can use a Cognitive Task Analysis, because the outsider knows so little. What the subject matter experts don't appreciate is how much of their own knowledge has become automatic, so that they aren't even aware of what they know. Their expertise isn't in the form of knowledge such as facts. It is in the form of practices, the way they see the job.

Example 8.2 Cognitive Task Analysis: Nurses in a Neonatal Intensive Care Unit

Crandall and Calderwood (1989) used Cognitive Task Analysis methods to find out how nurses working with microbabies (who weighed only a few pounds) could judge when the babies were developing infections and needed antibiotics. In many cases, the nurses were able to make these judgments before the blood tests showed any problem. With babies this small, getting started with antibiotics could be the difference for survival. When asked how they did it, the nurses couldn't explain. They said their skill was due to experience, or intuition, and left it at that. Using cognitive

probes with critical incidents, Crandall and Calderwood ferreted out the actual cues, some of which had never appeared in the medical or nursing literature.

At the conclusion of the study, the head of the Neonatal Intensive Care Unit requested that Crandall come back to teach these cues to the nurses on her staff, even though the data had come from these very same nurses. The nurses had not been able to articulate what they knew about diagnosing infection, because their skill was in how they saw the babies, not in the form of insights. Crandall and Calderwood were able to detect the subtle cues, and make them observable by others.

Sometimes, designers don't perform a Cognitive Task Analysis because they don't know how. For example, one team of designers took on the job of building a simulator. They hired a former operator for their team, someone who had worked on an earlier version of the equipment. Everyone on the team knew that the equipment had changed since then, but they never brought in a current user, because they didn't know what to ask. The next section describes some approaches they might have taken.

METHODS FOR PERFORMING COGNITIVE TASK ANALYSIS

There are a few methods that are in common use.[58]   Figure 6 presents four approaches: questionnaire/interviews, critical incidents, controlled observation, and analytical methods.

Interviews and Questionnaires are the standard data-gathering techniques used in many disciplines.   Interviews can range from unstructured to highly structured. Most task analyses use interviews, and can build on the questions to probe more about cognitive processes. What is unique about Cognitive Task Analysis is that the focus of the interviews or questionnaires is on the key decisions, cues,
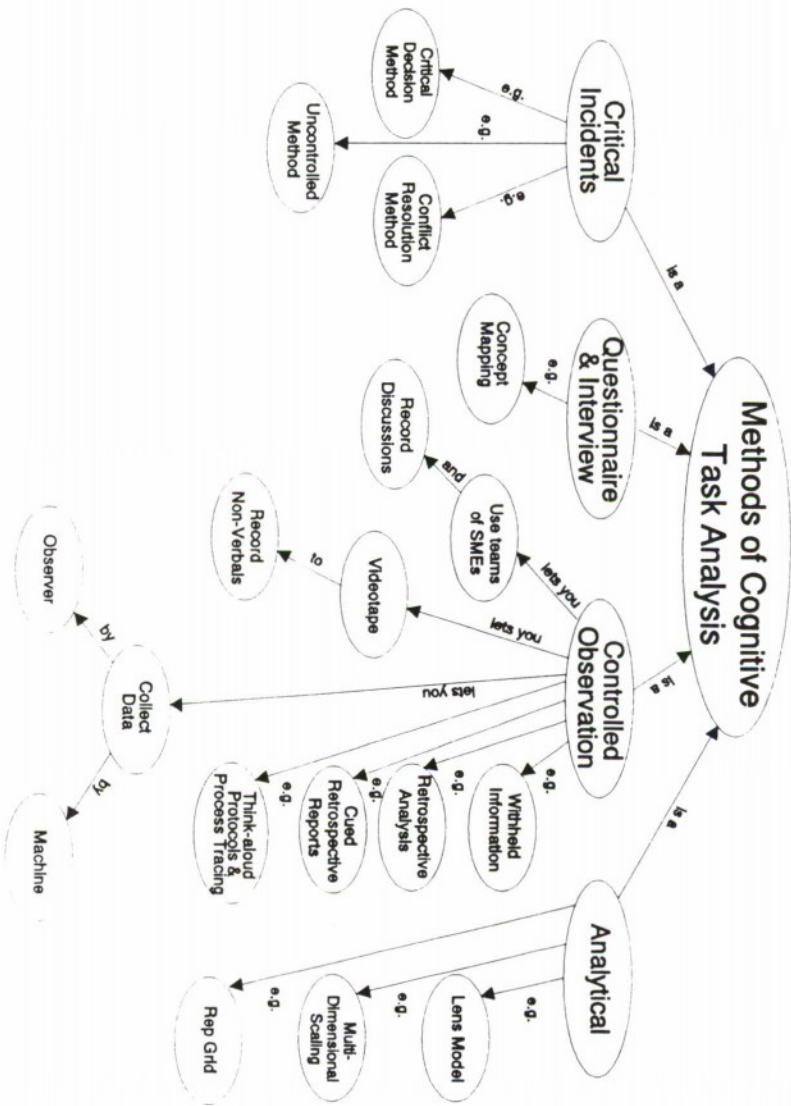
Figure 6. Concept map showing methods of cognitive task analysis.

distinctions, and so forth listed above. The intent going into the interview is to probe about these processes.

A list of cognitive probes is presented in Table 8. The list includes probing for cues, background knowledge, goals, situation assessment, options that were available, the basis for choosing among options—if that is what the person did—or the possible reasons for not having to choose, important experiences, aiding, and hypothetical scenarios. Depending on the project, different probes would be used, rather than the full set.

One type of semi-structured interview method is Concept Mapping,[59] in which the interviewee identifies the primary categories in a domain, and describes their relationships. Figure 6 is in the form of a concept map, showing the central kernel, Cognitive Task Analysis, the four types of Cognitive Task Analysis, and some information and methods of each. To perform concept mapping, you either ask the subject matter expert to describe the task or work environment, adding nodes and linkages, or you have the subject matter expert alone draw the concept map.

A second type of Cognitive Task Analysis shown in Figure 6 is the use of critical incidents. Critical incident methods usually consist of probing how people performed nonroutine tasks sometime in the past, during actual operations. Critical incidents are a very fruitful source of information. They are instances in which special expertise is needed, so they are opportunities to learn about some of the more subtle aspects of expertise. In routine cases, people perform many actions without thinking, whereas with novel tasks there is little expertise to begin with. Critical incidents fall in-between. There are demands on a person's abilities, and they are memorable so a person can recall with fair accuracy many details of the event. One critical incident probe is the conflict resolution method,[60] in which the person recounts the incident, and at different points in the episode is asked to imagine that the particular outcome had not occurred, and to hypothesize what else might have happened, and for what reason. This technique challenges the interviewee to come up with exceptions and special cases, rather than the stereotyped sequences of causes and events.

Table 8. Cognitive probes.

Probe Type	Probe Content
Cues	What were you seeing and hearing?
Knowledge	What information did you use in making this decision, and how was it obtained?
Goals	What were your specific goals at the time?
Situation Assessment	If you had to describe the situation to someone else at this point, how would you summarize it?
Options	What other courses of action were considered, or were available to you?
Basis of Choice	How was this option selected/other options rejected?
Experience	What specific training or experience was necessary or helpful in making this decision?
Aiding	If the decision was not the best, what training, knowledge, or information could have helped?
Hypotheticals	If a key feature of the situation has been different, what difference would it have made in your decision?

Another technique is the Critical Decision method.[61] Here, the person describes a nonroutine event requiring skill, and the interviewer goes through four cycles. In the first cycle, the interviewee briefly describes the event. In the second cycle, the interviewee pins the event to a timeline, as a baseline for checking consistency. In the third cycle, the interviewer uses some or all of the cognitive probes presented in Table 8, to understand the processes underlying the decisions. In the fourth cycle, the interviewee compares his or her own performance to that of a novice, pointing out where a novice might have misinterpreted events, missed cues, or made other mistakes.   The Critical Decision method was used to collect the incident accounts from which the RPD model was derived. It was also the method used in the preceding example of identifying cues for infection in microbabies.

In using cognitive probes to study critical decisions, you might want to modify the probes themselves. You could ask about the person's goals for that particular case, and contrast these to goals that a less-experienced person might have pursued. If a person selected an option or CoA without considering others, you could ask which options a novice might have considered, and what reasons a novice might use in selecting a poor option.

Controlled observation is another way of performing a Cognitive Task Analysis, by studying nonroutine incidents while observing performance and even questioning the decision maker during the task.[62] One approach is to have the person say out loud what he or she is thinking; this is referred to as collecting think-aloud protocols. Because think-aloud protocols can possibly interfere with performance, a variation is to have the person complete the task and then ask what went on—this is called retrospective analysis. However, the person may forget some details. A third variation is for the person to complete the task and then watch a videotape, or some other record of performance, and use this as an aid to remembering what had happened—this is called a cued retrospective report. Still another method is to withhold information needed to perform the task, and see what questions the person asks. Another variant is to have a team perform the task, to see what the members say to each other, and to collect these data less obtrusively than with a think-aloud protocol.

With controlled observation techniques you can videotape
performance, and thereby capture nonverbal actions. Also, the
technique allows data collection by an observer and also by a
computer, e.g., an analysis of a sequence of button pressing that might
be too quick for an observer, can be readily recorded by a
computer.[63] Smith et al.[64] performed an excellent example of
controlled observation in a study of aviation troubleshooting. They
presented qualified pilots with a written scenario of cues and events,
and asked the pilots what was wrong with the aircraft. By recording
the questions that the pilots asked, Smith et al. were able to construct
scripts that described the way each of the pilots understood the system
and the malfunction.

The fourth approach to Cognitive Task Analysis is to use analytical
methods. These are like little experiments used to pin down the cues
and dimensions used by subjects. Figure 6 lists three analytical
methods, the Brunswik/Lens model, Multidimensional Scaling, and the
Repertory Grid method. The Lens model[65] is a way to figure out
how much a decision maker relies on different sources of data, by
seeing how the judgments shift as the cues are changed. To do
Multidimensional Scaling[66] you would have a person make large
numbers of similarity judgments to measure how the person sees the
relationships between different stimuli. The result is to get a map of
the different concepts and their relationships. The Repertory Grid
method[67] asks people to tell how two objects are similar to each other
and different from a third; the answers people give show the types of
dimensions they use in their situation assessments.

There are other methods as well, such as field observation, which
can be used to help fill in each of the four methods shown in Figure
6.

When should you use each of these four methods? The interview
or questionnaire approach is the most general and common method.
It can be used by itself, or with most of the other methods. One
problem with using it by itself is that it usually elicits general and
idealized answers about how to perform a task, rather than cutting
through to details and contextual nuances. Therefore, it is not so
productive a method as the others. But in some settings, it is the only
feasible approach.

The critical incident method captures details and context, and provides a rich source of data. It seems to work best when the person being interviewed has a certain amount of expertise and personal experience. You should consider whether the domain allows feedback to the decision making. In situations where there is no feedback for actions, people may not develop expertise at the task. They also may have trouble recalling incidents because, without feedback, there really isn't a story to remember. The method also takes time, approximately 45 minutes to an hour to effectively probe an incident. The critical incident approach doesn't work well for tasks that are procedural, such as starting an engine, because people usually have trouble remembering critical incidents.

The method of controlled observations is also very useful. The advantage it has over the critical incident approaches is in being able to control and manipulate key features of a task, and in being able to present the same task to different people, e.g., both experts and novices. This method also avoids the problem of people being unable to recall any incidents because the procedure presents the incidents for them to handle. This method has several disadvantages. One is that it requires more up-front work to prepare the scenarios. Another disadvantage is that it is less likely to uncover new factors than the critical incident method, because the researchers building the scenario cannot know in advance what to include. Therefore, the method may be more useful for confirming hypotheses than for generating new ideas.

The use of analytical methods takes the most work to collect the data, and meets with the greatest resistance because the tasks can be tedious. Once collected, the data analysis is probably easiest with this approach. The analytical methods appeal most strongly to researchers who want to collect clean and unambiguous data that can be reported and published. The great advantage of these methods is that they do not rely on introspection, and so the data are more acceptable to those who have little tolerance for the ambiguities inherent in asking people to describe their own thought processes.

On the negative side, methods like Multidimensional Scaling are time consuming and not tremendously informative for applied purposes. The Lens model is impossible to use unless all the relevant

cues can be controlled and measured, so it is inappropriate for most
field settings. The Repertory Grid method is the most applicable of
this set, but the data it yields, sets of dimensions and categories, do
not go very far towards explaining how people make specific types of
decisions. Therefore, these methods are the least useful of the ones
described.

How can the findings of a Cognitive Task Analysis be packaged
and represented for design engineers? Otherwise, all the knowledge
will stay in the heads of the data-gatherers and none will make it into
the system. Thus far, several formats have been used for
representation. These include:

• Lists of critical cues and relationships used by experts and
missed by novices (Means & Gott, 1988)

• Annotated incident accounts, in which the incident is described
as a story, and the critical cues and relationships are highlighted as
annotations, so they can be understood in context (Crandall &
Calderwood, 1989)

• Diagrams of the judgments and decisions made during the
incident (Kaempf et al., 1992; Pew, Miller, & Feehrer, 1982)

• Formatted incident accounts, in which the events, inferences,
and decisions are laid out in parallel columns (Kaempf et al., 1992;
Pew et al.,1982)

• Scripts of the way users handle the task and make decisions
(Smith et al., 1988)

• Lists of the important types of decisions (Kaempf et al., 1992)

• Concept Maps (i.e., Figure 5) (McFarren, 1987; McNeese et
al., 1990)

• List of major concepts used in situation assessment (Miller,
Wolf, Thordsen, & Klein, 1992)

• Software programs modeling the information flow during
decision making (Woods, 1993; Zachary et al., 1991)

For any project, you may only want to use a few, or even just one,
of these representations. The goal is not to build fancy
representations, but to give the design team an idea of what the
operator is thinking about. Some of these representations are
embedded in the account of the critical incident or the scenario, so you
can put them in the context of doing the job. The purpose of these

representations is to enable you to gain a better understanding of the key decisions, cues, patterns of cues, and strategies that operators will use.

Because the purpose of Cognitive Task Analysis is to identify decision requirements, let us specify how this can happen. Table 9 lists some of the steps that could be taken to define decision requirements.

First, you would map out the task, using a task analysis. If no one had conducted a task analysis and time was short, you might rely on interviews with subject matter experts, and also use concept maps. The intent is to get an overall perspective on what the operator will be doing, and under what conditions.

Second, you would identify the decision points. These are the places in an incident where a person had to choose a diagnosis or a course of action, or had to update a situation assessment. It is a decision if alternatives existed that might have been chosen by someone with little experience, even if the subject matter experts never considered any alternatives. If you use a critical incident approach, these all fall out fairly easily. They also emerge during Concept Mapping and controlled observations. They would also come up during interviews, but the risk is that the interview will stay at too general a level. Subject matter experts have been known to take the interview as an opportunity to expound on their philosophy of the domain. To be meaningful, the decision points should be anchored to specific events as closely as possible.

Third, you can cluster the set of decision points, e.g., combining those regarding the adversary's intent, those regarding the weapon system to be used, and so forth. In this way, the generic categories are built up but are still linked to actual events, or to observations during simulated incidents.

The fourth step is to prioritize the decision points. You can put a lot of energy into this step, but the intent is fairly simple—pick out the most important decisions and spend your time on those, ignoring the trivial decisions.

Fifth, for the important decisions you would try to determine the decision type, e.g., was it a diagnostic decision or a course of action decision? What strategies were used, and what strategies might be

Table 9. Steps in the identification of decision requirements (Miller, Wolf, Thordsen, & Klein, 1992).

Map out the tasks

Identify decision points

Cluster the decision points

Prioritize the decision points

Identify the strategies and inferences for important decision points

used? How was the strategy carried out—what were the critical cues or patterns of cues? How were inferences drawn from the patterns of cues? There may be important cues that the operators did not mention because they weren't available, but which could be added to the new interface; these certainly should be included. You can set out different scenarios for how the operator might use the system in making decisions in the future, and try to imagine what could go wrong (e.g., faulty sensors, conflicting data) to anticipate what the operator would need to do. The process of filling in the expanded decision requirement already is getting you inside the heads of the operators, trying to think the way they would.

At this point, you would have a set of key decisions, along with the strategies, cues, and contingencies for each. This would constitute an expanded decision requirement that could serve as a baseline for the design process.

The use of decision requirements, and the representations of cognitive processing, can be applied to many purposes;[68] our concern in this report is on using NDM to support the system design process. For our purposes, Cognitive Task Analyses are important to identify the important decisions. Sometimes, it may not be possible to collect Cognitive Task Analyses, and the design team may hypothesize what the key decisions are, using a knowledge of the task and of NDM. In any event, the objective is to suggest ways of helping the operator make better decisions. The next chapter discusses ways of using decision requirements for system design.

DECISION REQUIREMENTS — THEIR ROLE IN
COGNITIVE SYSTEMS ENGINEERING

This chapter describes some ways you can use Naturalistic Decision Making during the design process. If you can identify decision requirements and use these to guide system design, the result should be systems and interfaces that work better, particularly during the tough parts of the job. Many factors go into system design, and decision requirements are only one thing to consider. The operator has to perform the basic parts of the job, as well as handling the difficult cases. The design has to keep workload from getting too high; it has to help the operator focus attention where needed; it has to stay within memory limits. It would be a mistake for you to make decision requirements the major driving force during the design and development process. But it can also be a mistake to ignore decision requirements, as happens too often. The premise of this chapter is that it is possible to incorporate decision requirements into the other considerations for generating a robust system.

The discipline of Cognitive Systems Engineering[69] has recently emerged to help design engineers take cognitive factors into account. If you are given the job of building a power source for a new computer, you would ask many detailed questions about the features of the computer, e.g., its size, peak energy requirements, most sensitive components, the environments in which it must run, and so forth. If you are building a system for an operator, you might need to ask similar questions about the features of the operators, e.g., their memory capability, their information-processing capability, their

ability to make discriminations, the strategies they use to derive inferences, and so forth.

Cognitive Systems Engineering is attempting to find ways to identify cognitive requirements and make them part of the system design. Functions such as memory, capacity limitations, attention, and decision making need to be considered.

- Systems that ignore memory limitations can leave the operators bewildered, flipping back and forth between different screens and trying to hold details from each to make critical comparisons.
- Systems that ignore capacity limitations pile on the information and create workload problems at the worst moments. This is particularly true of support systems that make easy parts of the task even easier, but make difficult parts of the task almost impossible. (An example is presented below of the Flight Management System of advanced aircraft.)
- Systems that ignore attentional limits can leave the operators helplessly scanning the screens trying to find what they need. An example is the current screen layout for the weapons director of the AWACS aircraft. During emergencies when an airplane is running out of fuel, the job of finding the closest tanker in a swarm of symbology can be overwhelming.
- Systems that ignore decision requirements can turn the operator into a spectator as events unfold more quickly than they can be controlled. The next section describes how decision requirements can be identified and used.

Decision requirements are important all through the design and development process. They are important during:

- Early concept development. Because the system must be designed to help operators make important decisions, you can identify these decisions at the very beginning. Decision requirements therefore become part of a front-end analysis.
- System specification. When you design the system itself, and particularly the human-computer interface and decision support systems, you can use decision requirements to ensure that operators will be able to achieve situation assessment, and make diagnosis decisions in sufficient time to stay ahead of the curve.

• Test and evaluation. In assessing whether the system does its job, you can include decision requirements among the criteria.

• System redesign. After the initial version has been fielded, you can identify decision requirements and incorporate them into retrofit and planned modification efforts.

EARLY CONCEPT DEVELOPMENT

The use of NDM may be most valuable during the early concept development stage. It is easiest to identify decision requirements for system redesign, because the system has already been fielded, and there are trained operators with experiences to relate for a Cognitive Task Analysis. However, decision requirements are best addressed during early concept development and system specification. True, Cognitive Task Analysis is much more difficult then, but so is every other design tool.

How can you analyze decision requirements before the system has been designed and delivered, and before there are users to interview? There are a number of steps you can take.

The Statement of Need should contain the initial understanding of major decisions, because the new system is intended to serve a function, and this function usually includes decision making. So, the system requirement is the starting point. The people who formulated the Statement of Need can be interviewed to pin down the problem more exactly, and even to collect the incidents that convinced people to go to the effort of requesting a new system. These incidents are another source of data. If possible, the people involved in the events can be interviewed to gather details, and a Cognitive Task Analysis can be performed on critical incidents.

Background interviews and data collection activities can define the information available to make decisions, particularly under difficult conditions. The data available in the field will constrain the types of decision strategies operators can use.

There are usually analogous systems to study, or, at least, people performing analogous sub-tasks, and Cognitive Task Analysis can be used to find out how they do their job.

Sometimes, an analyst may want to construct a scenario of how the decisions will be made.[70] During early concept development, it is difficult to envision the decision process, and there are many unknowns, so users and designers alike may keep their focus on a general level. They may realize there are gaps, but trust that they will be able to fill the gaps later on. Often they are right. But when they are wrong, the project comes to a halt. Sometimes the project team has to go back to the beginning, because they stumble over a problem that had been swept under the rug. If you want to anticipate decision requirements, you can lay out a scenario, with a reasonable sequence of events, and track the decisions. At each step, as the situation develops, you can anticipate what decisions will have to be made, what cues will be available, and what pattern of cues will be important. It is not easy to construct scenarios when so much information is missing, and sometimes you may have to make assumptions and/or lay out alternative scenarios. The payoff is that decision scenarios can make very clear the inconsistencies and leaps of faith. By getting down into the details, the step-by-step decisions, you can anticipate requirements that would otherwise be obscured and dismissed by a comment such as "Oh, we'll take care of that later."

Example 9.1 Using decision scenarios: JSTARS
The JSTARS aircraft, discussed in Example 1.3, was designed to perform a new function, looking over the edge of battle lines to see where the adversary was moving. One important feature of the aircraft was the Self-Defense Suite. The team performing a Cognitive Task Analysis to identify decision requirements for the Self-Defense Suite tried to construct scenarios of how the operator would make critical decisions such as judging when the aircraft was in danger. They kept running into disconnects, where the necessary information was not going to be available, or the key relationships were not going to be displayed, or the coordination for team decision making was not in place. These disconnects helped to alert the users that the aircraft was going to have trouble defending itself, which identified

additional decision requirements at both the individual and the team levels.

The example shows that you can spot inconsistencies even in the early design stages if you try to form a coherent story of how the primary decisions are going to be made in the context of the mission. All these activities should allow an analyst to infer the decision strategies that operators will use, and the key cues and relationships that will have to be highlighted in displays.

This example demonstrates that a decision requirements approach can be used for initial development, and avoids some of the pitfalls of rushing to build new software versions of the latest gizmos that excite the research and development teams.

SYSTEM SPECIFICATION

Once the basic design concepts are in place, the next step is to prepare specifications. We need to be clear about what the NDM framework can contribute to this phase, and what is still the designer's art. The decision requirements can show what needs to be displayed, but they cannot show how it should be displayed.

For example, returning to the displays in the AEGIS cruiser Combat Information Center, a conventional task analysis showed that the altitude of potential threats needed to be presented. The Cognitive Task Analyses showed that changes in altitude, and even changes in the rate of change of altitude, needed to be presented. The system users were adamant about wanting the altitude data shown on the primary displays, and not off to the side. But that was it. No analyses generated ideas about how to display the altitude data. This must remain a function of the creativity of the designer.

There is one way to assist designers in generating display concepts—prepare databases showing previous efforts to display the same types of information. These would resemble the pattern books used by architects, who find it helpful to skim through different ideas for implementing a given concept, such as an Early American Colonial

style house. It should be possible to assemble pattern books showing different ideas for presenting altitude data, or for presenting rate of change trends, or changes in the rate of change—acceleration as well as velocity. These types of interface and decision support pattern books would be a form of corporate experience.

During system specification, storyboards have been very helpful in evaluating design concepts.[71]   It would not be productive to use storyboards during early concept development, because more effort would go into the storyboard details that don't matter at that stage, than into the decision scenarios.   But for system specification, storyboards can help the operators and other members of the design team visualize what is going to be built.

The NDM framework could be used for appraising the specifications and the storyboards. The specifications could be run through the decision requirements, as through a filter, to see if all the important decisions can be made, given the information at hand, and under the conditions of intended use. This appraisal would find remaining gaps, and would increase confidence in the design.

The decision requirements could be used for appraisal from a negative rather than a positive perspective. Instead of trying to see if the specified system allows decision making, the appraisers can try to imagine ways in which decision making might break down, because of the system design.   Kyne, Militello, and Klein (1992) have described this as a Pre-Mortem technique, using mental simulation. Kyne et al. studied the way people evaluate their own plans, often glossing over weaknesses. The Pre-Mortem exercise is a way to help the planners or designers shift from being advocates for their own design, to the perspective of quality assurance. The designer reviews the system and then is told to imagine that, during operations, the system has fouled up—operators were unable to handle the decision requirements. But nothing more is known. The designer must think of different ways in which the system might have failed. This strategy helps to uncover flaws that depend on the context of activities, to show that under certain conditions users may run into trouble operating the system.

TEST AND EVALUATION

Too often, a system is built in a way that does not support decision making, but this is not noticed until T&E. In most cases, the testing during development is intended to find out if the software works as planned. Usually there isn't time to evaluate whether the operator can use the software to perform the task. For this, and many other reasons, T&E is a critical part of design.[72] If the testing criteria are prepared in time, they can be used during the system specification phase to guide the developers.

You can incorporate the NDM approach into T&E, to construct evaluation scenarios that stress the system, searching for ways to make the user/system crash. Developers expose hardware and software to such ruggedization tests. But, typically, developers just put the system and HCI through its paces to see if it works as planned, without trying to see what will make it break down. If the evaluators can identify critical decisions, they can incorporate these decisions into the testing scenarios, to see if the operators make poor decisions.

Example 9.2   Using test and evaluation to assess decision support: The watchstander station

Remember the watchstander example (Example 1.4). During an emergency requiring the activation of a rapid reaction team, the initial notification would go to a watchstander who needed to figure out which support organizations had not gotten the automated call-out message reporting the emergency. The watchstander had to find other ways of communicating with these organizations, or had to arrange back-up support. However, it appeared to evaluators that the screen design was going to get in the way of this decision. The screens showed the organizations that had been contacted, and those that had already responded, but it was impossible to figure out who had not responded, without switching back and forth between the different screens, and going through every entry to see if one was missing. The system developers would not have identified the problem

during user acceptance testing, because that testing was simply to determine if the system did all the things it was supposed to, without crashing. Moreover, typical T&E might not have uncovered the problem either, because the T&E scenario would likely have been straightforward. Because the evaluators were able to infer some of the key decision requirements, and to identify areas where the system failed to satisfy those requirements, they were able to design T&E scenarios that probed for system weaknesses. They recommended building a scenario where a key resource organization did not call back, to observe how long it took for the watchstander to notice this problem.

The preparation of test and evaluation scenarios to assess how well a system supports decision requirements is a new and interesting challenge for the design team. If these scenarios are constructed early enough, you may find that the designers start to make modifications in advance.

SYSTEM REDESIGN

This phase is the one where Cognitive Task Analyses can be used most fully. The system is in place, there are experienced operators, there are critical incidents to probe, so it should be possible to perform a thorough assessment of decision requirements to see where the original system needs to be improved.

The redesign effort can be relatively quick and inexpensive. That is because Cognitive Task Analysis does not have to perform an exhaustive compilation of information transfer, such as a data-flow diagram, or a complete inventory of necessary and available skills and knowledge, such as a conventional task analysis.

Example 9.3   A minimalist use of the NDM approach: Redesigning the AWACS weapons director interfaces

This project was intended to demonstrate Cognitive Systems Engineering. Klinger, Andriole, Militello, Adelman, Klein and Gomes (1993) selected the AWACS weapons director position for the demonstration in June 1991. The rationale for selecting this station was that high-fidelity simulators existed at Brooks AFB, so the redesigned interfaces could be carefully evaluated. The researchers had no prior familiarity with the AWACS weapons director job, and had no indications that the current design was inadequate in any way. Cognitive Task Analyses were performed during the summer and fall of 1991. These were Critical Decision method interviews with weapons directors. There were some delays in the interview schedule due to alerts because of continued tensions with the Iraqi government. (Desert Storm had concluded in February 1991.) Once the interviews were completed, analyses were performed during the fall of 1991. The key judgments of weapons directors were identified, particularly the judgments that were made difficult by limitations in the current interface. These judgments included locating high value assets (i.e., tankers) under time pressure, when an aircraft needed fuel immediately, and judging the relative speeds of friendly and adversary aircraft closing on an engagement. An additional set of cognitive requirements were identified that had to do with memory and attention problems. The researchers completed their redesign specifications and storyboards by January 1992. The software was completed by April 1992, and the redesigned system was evaluated from May through July 1992.

The evaluation showed that the revised screens were highly effective at bringing new weapons directors up to speed. The revised screens were compared to the current displays.

Operators who were experienced with the current displays (ranging between 266 hours-4300 hours) were given 4.5 hours of training on the revised screens. We asked an

experienced Weapons Director to perform a blind evaluation of the overall performance of 17 Weapons Directors who had used the current system in one testing session, and the revised system in another. The data showed that the rated performance with the current system was below average (3.77 on a five-point scale where 1 = high and 5 = low performance), whereas for the revised interface the overall rating was 2.82, better than average. The difference in ratings was statistically significant at the .05 level.

Looking at specific measures, the most important is whether the Weapons Director allowed hostile strikes to be completed. The revised interface reduced this number by 20%. It also reduced the average depth of penetrations by hostile aircraft by 30%. This not only means that fewer hostile aircraft were completing their missions, but also fewer hostile aircraft were even threatening friendly ground assets. This important outcome was achieved while friendly aircraft losses were decreased by 15%. So, fewer friendlies were shot down, fewer hostiles completed their missions, and the hostiles were not penetrating the friendly airspace as often with the revised displays. With the new displays the operators were able to increase air refueling by 76% during simulated missions and reduce the aircraft returning to base for refueling by 18%. Moreover, with the current system 12 out of 17 operators had hostile strikes completed against them. With the new system only 9 out of 17 operators had hostile strikes completed. For the new AWACS weapons directors, the improvement in performance with the revised displays would have been difficult or impossible to achieve using a hardware solution alone. Just speeding up the data processing, adding more power to the work station, could not have yielded the impact of Cognitive Systems Engineering to speed up the users' decision making.

The AWACS project and others like it show the value of cognitive systems engineering for cost-effective upgrades of decision-making performance.

Example 9.4 Directly linking system design to cognitive task analysis: Anti-air warfare operations in the AEGIS cruiser

As described earlier, the purpose of this project[73] was to evaluate the use of NDM for designing human-computer interfaces and decision support systems.   Kaempf et al. conducted 17   Critical Decision method interviews with commanders and Tactical Action Officers and Anti-Air Warfare Coordinators of AEGIS cruisers.   Table 10 lists the primary decision requirements that were identified (Miller, Wolf, Thordsen, & Klein, 1992).

Because decision requirements are so important in this process, let us spend a little time going over the way the requirements in Table 10 were identified. They all came from the critical incident interviews.   For each interview, the important decisions were noted. Kaempf et al. prepared a master list of these decisions and determined which types of decisions were seen again and again.   These are the ones listed above. Each of them appeared in at least one incident, often in many incidents, and each can be traced back to the specific incidents from which they were derived.

The next step was to show, for individual cases, the transitions between decisions. Figure 7 shows a schematic for the decisions in different incidents. Thus, Incident 1 began with the decision A, to determine intent. You can look back to Table 10 and verify that the first decision requirement was to determine intent.   Incident 1 shifts to decision B, determining if it is a potentially threatening situation, problem, and so forth.

Once this was done, the next step was done by Miller et al. (1992), who put together all the instances of a decision requirement, to see what common cues and relationships were needed.   For example, there were seven instances of

Table 10. Primary decision identified in the critical incidents.

<u>CODE</u>

A. Determine intent: CIC crew attempts to determine the intentions of a track, such as whether or not the track is hostile.
B. Recognition of a problem: crew tries to determine if they are faced with a potentially threatening situation.
C. Take actions to avoid escalation: crew takes deliberate steps to avoid the escalation of an incident.
D. Take actions toward engaging track(s): crew takes preparatory steps needed to engage a track.
E. Monitor ongoing situation: the CIC crew monitors a situation to detect any changes in the situation.
F. Identify track: crew attempts to determine the identity (e.g., country of origin) of a track.
G. Allocate resources: the CIC crew attempts to allocate limited resources to deal with the current situation.
H. Prepare self-defense: crew takes steps toward self-defense, such as bringing up the CIWS.
I. Conduct all-out engagement: crew actively engages a track with a weapon system.
J. Monitor tracks of interest: crew monitors a track which has some significance to the current situation.
K. Reset resources: the crew returns ship resources to pre-incident status.
L. Collect intelligence: CIC crew actively tries to collect information on a track.
M. Trouble-shoot: crew tries to trouble shoot a system
N. Determine location: CIC crew attempts to determine the location of a reported track.
O. Other: goals not coded in the above list.
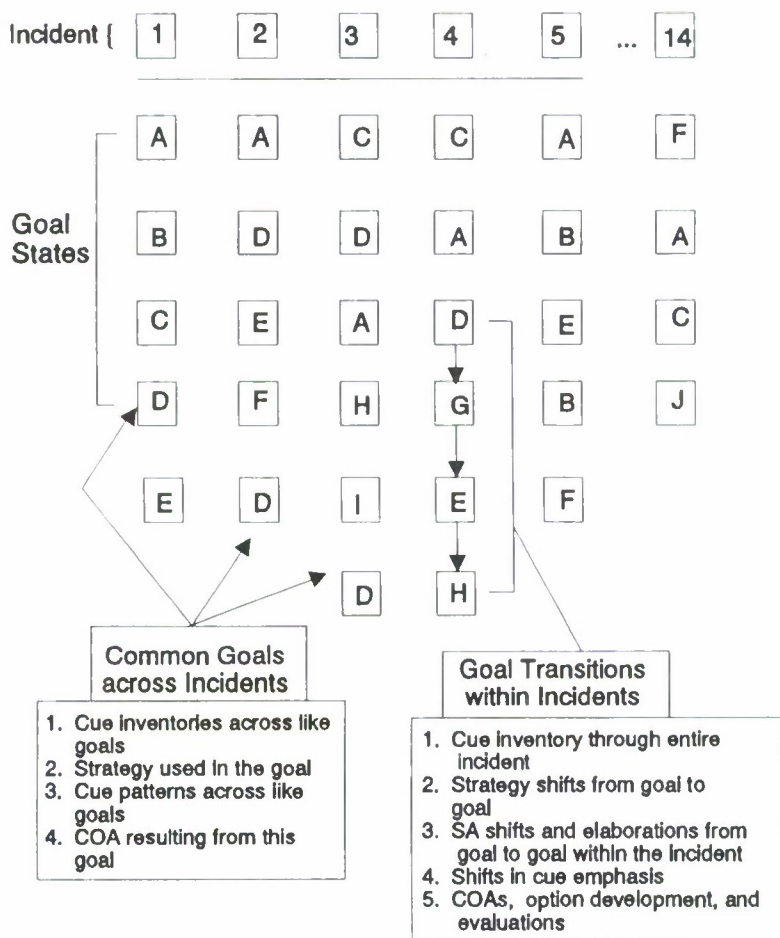
Figure 7.   Schematic.

identifying the intent of a track, derived from five different incidents. For these seven cases, Miller et al. compiled all the cues and relationships and inferences that were used, and made these part of the expanded decision requirement. These are shown in Table 11. You can see that crew members relied on intelligence reports, instances of recent hostilities, recency of the track, the course direction and profile, the range of the track and its origin, whether the range was decreasing, and so on to infer the intent. Notice that the last cue, ascent or descent, refers to simple direction, and not to sudden changes in ascent or descent. Why didn't the officers use changes in ascent/descent rates? Possibly it is not a helpful cue to them, or perhaps they were not able to make these judgments from the AEGIS displays. We need to consider subtle cues that may not appear in the record, so that we are not restricted by the limitations of the previous interface concepts.

For a designer, then, the decision requirement includes the key items of information. And all are linked to specific incidents, to convey how the information is used in context. In this way, the Cognitive Task Analysis leads directly to the expanded decision requirement.

Some of the concepts for the display are shown in Figures 8-11. The operator views the track, in Figure 8. He can call up a list of critical features, as in Figure 9. He can call up a historical record of altitude, as in Figure 10, or range, as in Figure 11. Figure 11 shows that the circles have been steadily shifting, and the range at the closest point has been steadily decreasing. During the actual incident, the commander suspected this was happening, but had no way of confirming the suspicion.

This example shows how decision requirements are identified and expanded to include the necessary cues. These cues can then be incorporated into the screen design so the operator has the necessary information readily at hand. Note again that these display concepts were not derived directly from the decision requirements. The NDM

Table 11. Expansion of decision requirement. This Table shows the type of information that was used to infer intent, in the seven instances where it was studied.

Cue

Intelligence

Recent hostilities/activities

New track

Course intercept, erratic, circling

Range

Point of origin

Change in range

Electronic Warfare bearing

Change in course

Knowledge of enemy tactics/weapon

Response to warnings (none)

Speed

Change in speed

Number of tracks

Altitude

IFF (Identify Friend or Foe)

Formation

Flight profiles

Vertical Air Speed (Ascent/Descent)
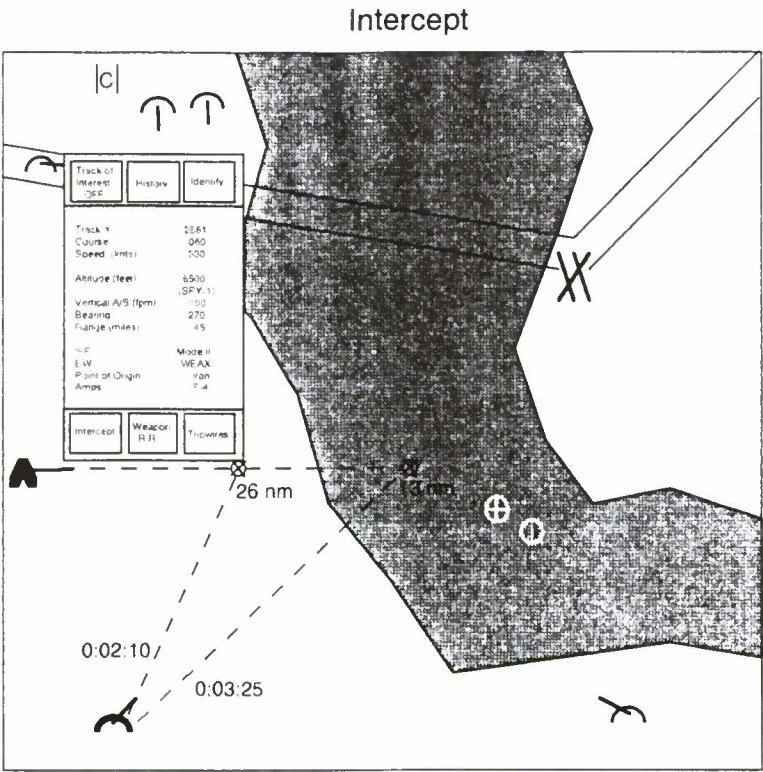
Intercept

Figure 8. Screen 1 Symbology.

Track of Interest	History	Identify

Track #	2123
Course	060
Speed (knts)	330
Altitude (feet)	6500
Vertical A/S (fpm)	100
Bearing	270
Range (miles)	45
IFF	None
EW	None
Point of Origin	Iran
Amps	F-4

Intercept	Weapon R.R.	Tripwires

Figure 9.   Screen 2 Track Information.
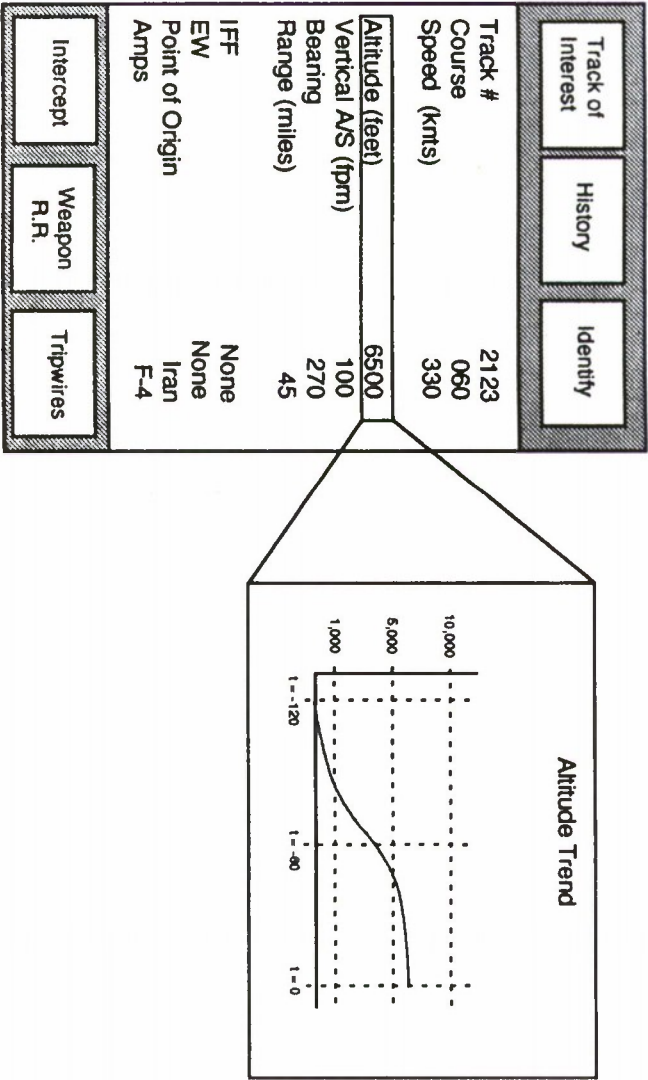
Figure 10. Screen 3 Altitude Trend.
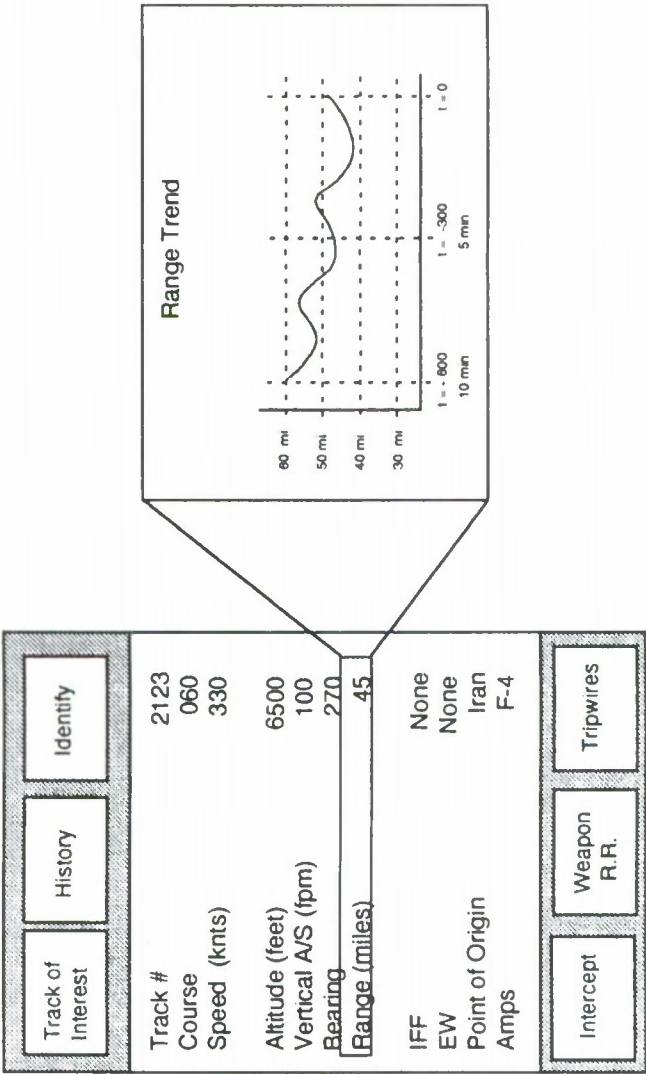
Figure 11. Screen 4 Range Trend.

approach is not a short-cut for the designer's skills and creativity. Rather, it provides the designer with additional information and ideas to help in specifying system features. The interface features were identified by carefully going through the critical incidents, suggesting display features that might have helped the crew at each decision point, and then combining the different features that were suggested. The final set of features grew out of the decision requirements for actual incidents.

INTERFACE CONCEPTS DEVELOPED TO SUPPORT NATURALISTIC DECISION MAKING

The NDM approach may give rise to its own set of interface concepts. In this section we examine some of the ideas that have been presented along these lines.

David Noble has designed a decision support system that performs feature matching to help diagnose situations. Noble refers to this decision support system as an RPD tool. Thus far, it has been used to support intelligence analyses and to alert crew members on an AEGIS cruiser that a potentially hostile aircraft needed to be monitored more carefully.

There are other decision support systems that are compatible with NDM, although they were not developed expressly for the purpose of supporting specific NDM models. For example, a system developed by Hair (1992, personal communication) is designed to help people make diagnosis decisions by evaluating the plausibility of different hypotheses, and to help operators keep track of data that are inconsistent with each hypothesis. We can also include case-based reasoning systems as being ways to support analogical reasoning. Riesbeck and Schank (1989) have written a comprehensive description of how to build systems that use analogical reasoning to derive inferences. A recent program illustrates that such systems may be ready for operational use.

Example 9.5   Decision support system that uses analogical reasoning: Making bids for manufacturing

Reed and Klinger (1991) designed and developed a system that uses analogues to help a manufacturing company generate and evaluate bids.   Previously, the company would receive requests to bid on making new parts.   The process of coming up with a bid was time consuming, and the accuracy of the bids was disappointing.   The analogical reasoning system collected previous bids in a database and used algorithms to enable the bidders to find similar cases.   The similar cases showed what the parts had actually cost, so the bidder had some idea of what cost figures to use.   The case history let the bidder see how to adjust the previous costs to meet the conditions of the new part.   The case history also described the process of manufacturing the previous part, so the bidder could envision a plan for making the new part or evaluate a plan in light of the earlier experience.

The bidding support system was designed specifically to demonstrate how case-based reasoning could be applied to current needs in the manufacturing domain. The system is being used and has reduced the time needed to bid on new parts.

Attempts to develop systems to support NDM will be primarily research efforts, although Noble's RPD support system is being developed for field use, and some of the case-based reasoning systems have also been fielded. The strongest use of the NDM approach is to strengthen the system development cycle, particularly as a front-end analysis technique. This chapter has shown how you could identify and expand the decision requirements and use them in the design process.

TEAM DECISION MAKING AND SYSTEM DESIGN

Looking back at the list of characteristics of NDM, in Table 2, one item that had to be taken into account was multiple operators, working in a team. The examples used in this report illustrate the importance of teams. The JSTARS self-defense suite operator has to work with the pilot and the mission control coordinator to arrive at critical decisions about changing course when faced with a threat. In the Vincennes shoot-down incident, the Commanding Officer made his decision to engage based on information he received from crew members in the Combat Information Center. The AWACS Weapons Director works as a team member with other Weapons Directors, and also with the pilots of the aircraft being controlled. To provide support for naturalistic decision making, we must understand how teams make decisions.

This chapter begins by describing several current accounts of team decision making. Then the issue of stress and decision making is revisited at the team level. Next, we discuss some implications of team decision making for evaluating decision support systems. The chapter concludes with a discussion of design teams.

A COGNITIVE FRAMEWORK FOR TEAM DECISION MAKING

The topic of team decision making has been receiving increased attention, and there are a number of insightful reviews for a reader who wishes to delve more deeply into the subject. These include recent chapters by Duffy (1990) and by Orasanu and Salas (1993), a

literature review by Dyer (1984), a report by Eddy (1989) on measures of team performance, a recent book on team performance by Swezey and Salas (1992), technical reports by Crumley (1989, 1990), monographs by Olmstead (1990), Orasanu (1990), and Kahan, Worley, and Stasz (1989).   This section will not try to synthesize all this material, but instead will briefly describe a framework that can be used to think about teams.

The idea is to take what we know about the way individuals think, and bump it up one level as a model of teams. A cognitive model of team decision making views a team as an intelligent entity, subject to all the cognitive limitations of an individual—limited memory, limited attention, limited situation assessment capabilities, and so on. The intent of the cognitive framework is to focus attention on the team, rather than on the team members, and to take advantage of our knowledge of individuals to better understand team decision making.

Morgan (1986) was one of the first to show that teams could be treated as cognitive entities, and Wegner (1987) discussed the phenomenon of a team memory: because different team members know different things, a team has the same job of figuring out how to retrieve information as does an individual. Thordsen and Klein (1989) showed that Cognitive Task Analyses could be performed with teams; because a team that is performing a task is automatically generating a think-aloud protocol, an observer can tell what the "team mind" is thinking about as well as any team member can, and without disrupting the activities.   Klein and Thordsen[74] showed that the deliberations of a team that is planning an action seem identical to those of an individual using mental simulation, and that teams are no more likely to make decisions by contrasting options than are individuals. Cannon-Bowers, Salas, and Converse (1990) showed that the concept of a mental model, which has been used to explain thinking and decision making of individuals, also applies at the team level, and that it is possible to study the way a team develops and uses a shared mental model of a task.

We see here the emergence of a view of team decision making that is akin to a cognitive process. For a human, memories are scattered through various parts of the brain, and linkages between events and ideas occur in parallel, without any central control. For a team, each

member brings his or her own experience, and no one can orchestrate the way different ideas are presented and combined because no one knows what is in the team members' heads. New ideas are generated that no single member would have considered, and team members adjust to unexpected events to keep the team on course for its objectives.

Figures 12 and 13 (from Zsambok, Klein, Kyne, & Klinger, 1992) are schematics which show how a team mind develops. The identity of a team strengthens to the point where members, who have just been doing their jobs, shift to trying to compensate for each other, to accomplish the team's goals. The team shows greater conceptual ability as it learns to make better use of the ideas and experiences of its members. The team achieves better control over its thinking as it learns to monitor itself and adjust to problems.

Figure 13 depicts some of the behavioral markers that have been found to be key indicators of effective teams.

• For strengthening its identity, an observer can see whether the team defines roles and functions, so that everyone has a clear responsibility, and knows what to expect from each of the others. Effective teams avoid micromanagement; teamwork requires that the leader perform specific functions rather than taking over the job of subordinates. You can gauge the extent to which a team achieves identity by noting whether members compensate for others who are having trouble. The fourth marker is whether any team members have become disengaged, and are allowed to drift away.

• For judging whether a team is expanding its conceptual level, you can readily detect whether a team seeks divergent ideas, by noting whether or not members are asking for different opinions or showing impatience when they are expressed. You can also assess whether the team tries to converge on a situation assessment, perhaps by reviewing its current state of understanding, or through the use of maps and diagrams. You could also note whether the leader explicitly makes the effort to seek synthesis and agreement (convergence). Another marker of expanding conceptual level is whether a team notices that it is missing information, or that there are different interpretations of what is going on; effective teams pick up on these and try to acquire critical data or clarify their interpretation. To judge whether a team is looking
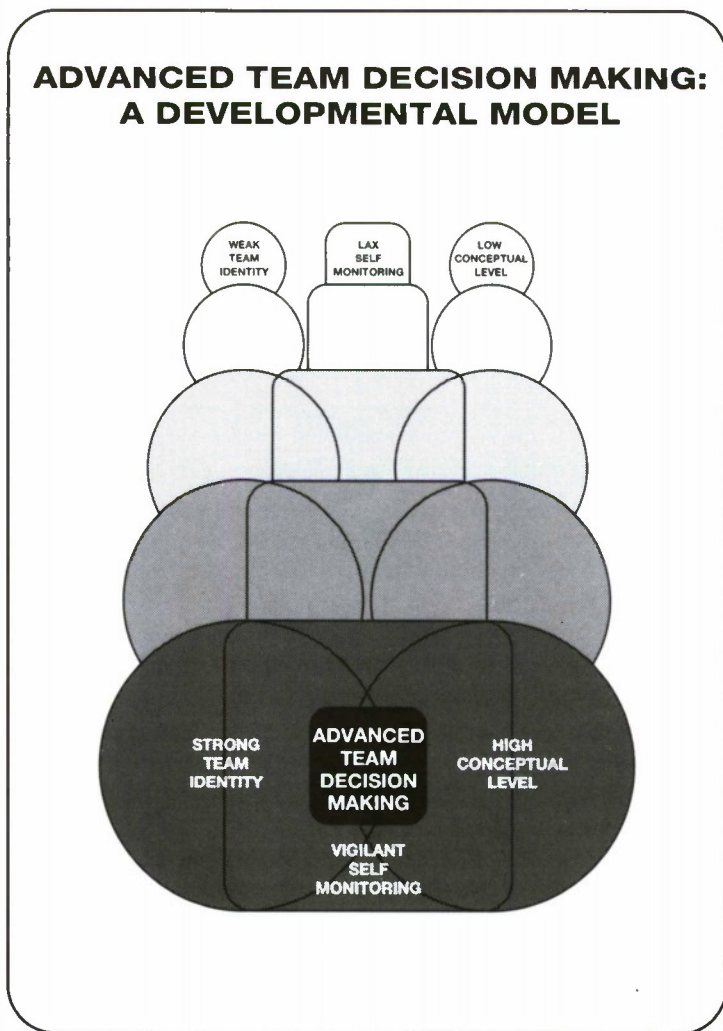
Figure 12.   Advanced Team Decision Making: A developmental model.

far enough into the future, you can keep a running tally of the time frame of each of the topics discussed. You can also tell whether the team members can envision the goals by noting how carefully the goals are described, how much detail is presented, and whether effort is spent finding out if anyone is confused or unclear.

• For assessing improvement in self-monitoring, you could watch how the team manages its time—does the team periodically assess the rate of progress, the tasks remaining, and the time needed to finish each task? Finally, you can watch to see if the team tries to do anything about problems it may be having.

The schematics in Figures 12 and 13 are intended to serve as a reference point for the topics discussed in this chapter.

TEAM DECISION REQUIREMENTS

During the phase of preparing system specifications, designers may find it useful to identify the different work teams responsible for various aspects of the mission, to see how the design of individual operator stations may affect the team decision making. The behavioral markers presented in Figure 13 are a checklist of possible system impacts. Each of these can be enhanced by a well integrated system, or degraded by a poorly thought-out system.

The team's identity can be affected by the nature of the design. Identity can suffer, disrupted by a system that lets any member communicate with any other, with no audit trail of who knows what. With this uncontrolled flow of communications, the role of the leader is easily compromised (see Duffy, 1993), and the potential for confusion, redundancy, and poor coordination increases. Systems can also foster micromanagement, by permitting leaders to review the outputs of each member. Engagement of team members can diminish as the system increases the isolation of each person. A well designed system can also bolster team identity, by making information readily available to members who might otherwise have been out of the loop. Compensation for difficulties can become easier with interchangeable work stations.
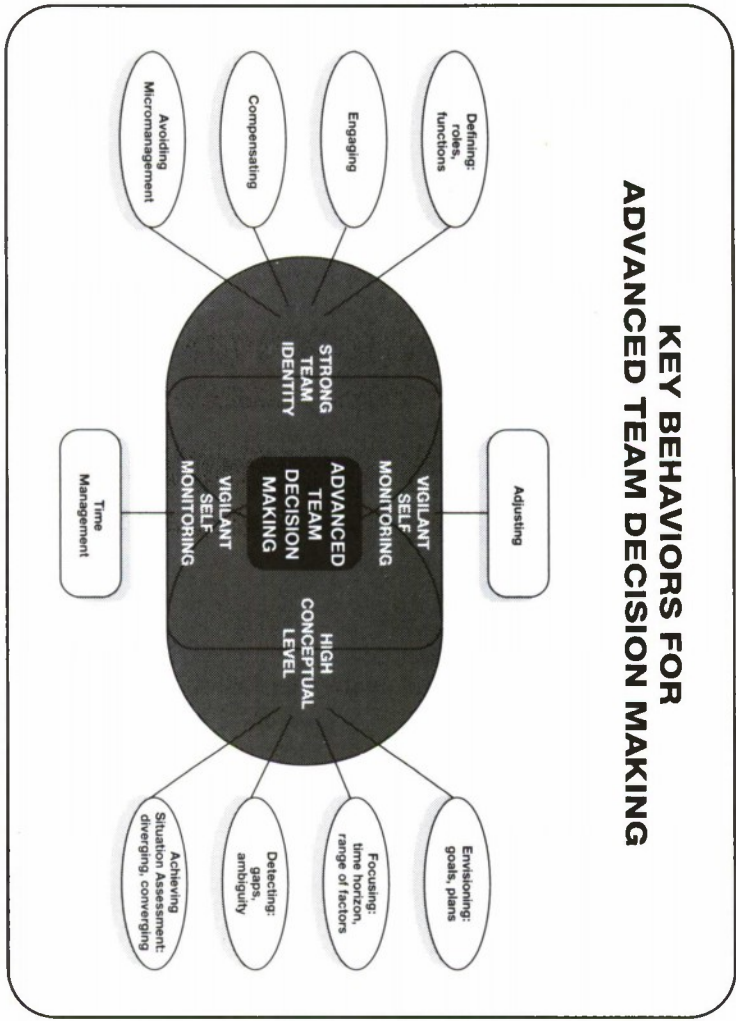
Figure 13. Key behaviors for advanced team decision making.

The team's conceptual level can be boosted by a system that lets each member check on the current situation assessment, presented on a common display. However, problems can arise due to information gaps and ambiguities. Because gaps and ambiguities are often signaled by nonverbal cues during face-to-face communication, the use of separate work stations may mislead the team into having more confidence in its situation assessment than is justified. Divergent ideas can be generated and rapidly disseminated by various types of groupware (hardware and software to facilitate group interactions), or they can be blocked by the reduced opportunity to pursue a line of thought across different team members. There are ways to support teams in keeping appropriate time horizons, but the lack of face-to-face communication can also lull the team members into believing that someone else is taking the broader perspective when, in fact, no one is.

The team's self-monitoring can be improved by a system that permits individuals to track what is happening to each other, and to the team as a whole, and to help the team synchronize schedules to better manage time. Adjustments can also be streamlined by features that support rapid reconfiguration.

Example 10.1   Disrupted team decision making: Advanced helicopter displays

Leon Segal (1989), a former helicopter pilot for the Israeli Defense Forces, has raised concerns about new displays. For the pilot, these displays are usually a great improvement. However, the displays eliminate cues that are important for team decision making. With the original, mechanical displays, a navigator knew what the pilot was doing at all times. The pilot's gaze would be directed at certain instruments while flipping on certain switches, or turning certain dials. Everything was out in the open, and coordination was maintained.

Some of the new displays make extensive use of CRTs. The screens are reconfigurable, and the buttons serve different functions, depending on the mode selected. As a result, the

navigator has much more trouble figuring out what the pilot is intending. The pilot/navigator team has much more trouble converging on a situation assessment. Coordination suffers, and, as Cannon-Bowers et al. (1990) would put it, there is a loss of a shared mental model.

The example shows how risky it can be to use conventional task analyses to improve work stations, without taking team coordination into account, and without understanding the cues that team members use to sustain their coordination.

The purpose of this discussion was to present some ideas about how system design affects team decision making. The development of groupware is an important topic in its own right. But even if a designer wants to work solely on an individual work station, the requirements will often include concerns about how the crew member interacts with others. The team decision requirements have to be identified and represented so the design engineer has some basis for taking them into account.

STRESS AND TEAM DECISION MAKING

The acute stressors that affect individual decision making may have an even stronger effect on the team, because they would disrupt the team interactions as well as the performance of the individuals. Further, team interactions may be more vulnerable to stress effects. Individual decision makers often do very well under stress, and in many studies, stressors such as noise and time pressure have little impact. In contrast,

• Time pressure can throw off the coordination among team members. Individuals may be able to use recognitional decision strategies to avoid the problem of time constraint, but teams do not have such shortcuts available.

• Ambiguity creates a cascading problem for teams, because no member can be sure of understanding the way the others are

interpreting the events, in addition to the uncertainty the individuals feel themselves.

• Noise can seriously degrade team coordination by preventing teams from using typical (i.e., verbal) communication pathways.

• Public scrutiny of performance may actually be less stressful for teams, because the failures of team members may be masked. Conversely, team members who feel responsible for the outcome may experience more frustration because they are in less control than if they performed the task by themselves.

• High workload poses a different problem for teams, because in addition to enduring the workload, they have to cope with the coordination difficulties when tasks aren't completed on time, and with the need to reallocate workload in the middle of the task.

System developers may find it useful to think about what stress is going to do to the team coordination needed to carry out a mission. By looking at the effects of stress we can deepen our understanding of team decision making.

EVALUATING DECISION SUPPORT SYSTEMS

Design engineers can use the cognitive model of team decision making to appraise the decision support systems themselves. These advanced systems, often using intelligent technology, can be considered to be team members.[75] By using the behavioral markers shown in Figure 13, designers can evaluate whether a decision support system is going to be a good team member or not—whether it will help the team move forward on the ten dimensions, or will hold the team back.

Example 10.2 A decision support system that is a poor team member: The flight management system of new commercial airplanes

For aircraft such as the Boeing 757 and 767, the Airbus 300A, and others, an advanced decision support system has been added to the cockpit. The Flight Management System is

a major step up from autopilot, and has the capability to direct the airplane's course and altitude throughout the flight. The goal is improved performance, due to reduced workload. The result is not always successful. During the routine parts of the mission, the Flight Management System does reduce workload. Unfortunately, these are already slow times in the cockpit, so the reduction doesn't help and makes tedious tasks even more boring. During nonroutine parts of the mission, the system can get in the way. A number of researchers[76] have reported that the Flight Management System is difficult to re-program whenever there has to be a sudden change in plans, especially during landing when Air Traffic Control redirects an airplane. Even in the last generation of cockpits, pilots have a difficult time adjusting to last minute changes requested by ATC, but the Flight Management Systems make the job even harder. Pilots enter in the new flight plans, but often are uncertain what the system knows. As Wiener et al. (1991) put it, the most common pilot reactions are (a) what is the system doing, (b) why is it doing that, and (c) what is it going to do next?

Some airlines have figured out how to handle the problem—they suggest that their pilots turn off the Flight Management System during nonroutine incidents.

In terms of team decision making, we would have a low tolerance for a team member who acted this way—unpredictable, impossible to read, resistant to redirection, and unconscious of the needs of others. Looking back at Figure 13, we would say that the Flight Management System added to role confusion, prevented compensation, tried to disengage the pilot, prevented convergence on situation assessment, hid information gaps and ambiguities, blocked divergent approaches, restricted the time horizon, and obscured the current goals. Time management and adjustment became difficult as long as the system was left on.

These shortcomings also suggest features that need to be added to the Flight Management System to make it a better team member.

The idea of evaluating a system as if it were a team member may seem unusual. The rationale is that advanced systems serve many of the functions of team members, without being accountable. As long as the systems work reliably and interpret and apply information as specified, they have been judged successful. You will be able to hold advanced systems to higher standards of performance if you have clearer expectations of how these systems must interact with the rest of the team.

DESIGN TEAMS

We have been looking at teams of operators, and at operator/system teams. Now we switch to look at the design teams themselves.

For virtually every system development, it is necessary to assemble a team of designers and engineers. Sometimes users will be included on the team. While this is an efficient way to get systems built, it is not always successful in building systems that do the job. Design engineers play the central role in most system development teams. Yet in most cases they aren't given critical information about what are the decision requirements, and how the operators will be likely to make decisions using the system. The design engineers may be provided with data flow diagrams and task analyses, but not with the decision flow diagrams and Cognitive Task Analyses they need to understand what the operator will be trying to do during critical incidents.

Systems are often constructed under great time pressure, and some developers, eager to begin work, may be unwilling or unable to find out what their system was supposed to actually do. But experienced designers have learned that you pay at the end for impatience at the beginning. For the design engineers, the barrier has been that the methods for identifying decision requirements have not been available, until very recently.

A makeshift solution to the lack of tools for identifying decision requirements has been to rely on the users, the decision makers who need the new system. A common mistake that system developers

make is to believe what the users say they want.   Users often misunderstand what the new system will do for them, or how it will change their job. They cannot visualize the end product the way the system developer can.   Few users can understand the subtle implications of a system design. The function of the user is to communicate the nature of the problem, not to specify the solution. Many design efforts fail because the developers try to build what the user has asked for, rather than trying to figure out what the user needs. It is crucial to listen to the users' ideas and needs, and users have an essential role on the design team. The point is that the user may not know what the solution is.

One possibility is to designate someone on the design team to be the operators' advocate—to represent the needs of the operators who will be eventually using the system, by helping the current users imagine how they will be making decisions in the future, and by helping the designer to imagine how the system will support that decision making. This is a function that cannot easily be performed by the current users, since they may lack the technical background. It may also be a challenge for the system developers, who may want to be provided the information about decision requirements without performing these analyses themselves.

Someone also has to be an advocate for the eventual operator when the development process is divided into modules due to complexity or time pressure. Often, different groups will work on different modules in parallel to speed the process up. This is efficient in terms of time, but it creates the major headache of making sure the different modules will work together.   Even if the developers assemble a system integration group, there is a danger that they will fixate on hardware and software integration, rather than decision integration.   It is necessary that the entire system support the decision requirements of the operators, but as the different modules are being assembled, both users and developers lose track of the decision integration.   For example, one management information system set up different teams, organized by the type of function to be built, and all through the design stage no one was sure how any specific user would move back and forth between the modules.

The reason for assembling a design team is to overcome the limitations of each of the team members. Well functioning design teams show the same characteristics as any other team—the ones presented in Figure 13. There is nothing new about extolling the virtues of a design team. One of the objectives of this report is to explain how a design team can identify and incorporate decision requirements into the design process.

The NDM approach applies to teams as well as to individuals. This chapter has explored the various facets of team decision making. Relying on a common perspective, a model of a team as an intelligent entity, we have considered the way that a design needs to take team decision requirements into account; we have covered the effect of stress on team decision making; we have defined an advanced decision support system as a team member to see the implications for design and evaluation; and we have examined the functions of the design team itself.

11

CONCLUSIONS

With the emergence of the field of Naturalistic Decision Making, we may be in a position to actively take the operator's needs into account. It is one thing to advocate for a user-centered system; it is another to construct the strategies for carrying out such an agenda. The current approaches to Cognitive Systems Engineering are all trying to specify ways of incorporating processes such as workload, memory, and attention into design. This report attempts to contribute to this movement by showing how to identify and apply decision requirements during design.

Decision requirements are different from performance objectives. For an operator of a self-defense suite aboard an aerial reconnaissance plane, a task may be to make sure that the aircraft is not exposed to unacceptable danger. A decision might be to judge whether a specific threat can attack the aircraft before friendly interceptors can intervene. A decision requirement is to make sure the operator can monitor the edge of the threat envelope (e.g., perceive time available for defense, compared to time needed for nearest interceptor to intervene), and can take course changes into account, as well as noticing immediately if the situation changes. Task analyses are concerned with the criteria for determining if a procedure was carried out. Decision requirements are concerned with understanding the way an operator will carry out the task—the specific diagnoses and action decisions along with the way the operator will derive inferences from patterns of cues. The contribution of NDM is to show how to derive decision requirements.

What will success look like? Success means different things to the operator and to the designer. From the perspective of the user, if we learn how to use decision requirements:

- You will see operators who can rapidly assess situations and can easily reassess them following interruptions or dynamic shifts in events.

- You will see operators diagnosing problems without getting confused or disoriented.

- You will see operators smoothly calling up new information to make diagnoses rather than feeling overwhelmed and blindly following along an unsuccessful path because it is too much trouble to start anew.

- You will see operators choosing more complex and context-specific reactions rather than sticking to stereotyped responses because there isn't enough time to do it right.

- You will see operators directly perceiving critical relationships, just through eye movements, rather than frantically paging back and forth between different screens.

- You will see operators decreasing their reaction time because they are able to detect key changes right away, just as sports car drivers have a better feel for the road and are able to carry out more difficult maneuvers.

- You will see operators giving themselves more time to gather diagnostic information because they are in better control of events.

What will system developer success look like from the perspective of the designer? If we learn how to use decision requirements:

- You will see designers using decision requirements to help organize the system features and the human-computer interface.

- You will see designers able to picture how the operator will use the system, especially during nonroutine incidents, the way an architect can visualize how people will live in a new house.

- You will see designers calling up analogues to figure out how to represent complex relationships.

- You will see designers anticipating problems that might be caused by a configuration.

• You will see designers identifying more problems during the early phases of system development rather than having to wait until Test and Evaluation (T&E).

• You will see designers who want to know the key decisions that operators make, along with the primary cues and relationships, as part of the background materials used during early concept development.

The NDM approach is intended to provide ideas and insights, not formal procedures. It will not develop into a lock-step procedure, the way task analysis and Instructional Systems Development have. The design process is too fluid, too context specific, for that to happen. The goal is a more modest one, to enable developers to take into account the more subtle aspects of human performance.

RECOMMENDATIONS

To assist systems developers who wish to use aspects of a NDM approach, this chapter presents the following recommendations:

1. You can request that decision requirements be identified and represented. We have discussed procedures for defining decision requirements at various stages in the design process.

2. You can request that the decision requirements be context specific. Furthermore, factors such as stressors can be considered as part of the context. The expanded decision requirements must fit the operators, the task, the domain, and the equipment. It is not enough to tell you that the operators will use a feature-matching strategy for diagnosis, or will rely on story building. That is just the beginning of the work, not the product. You can expect that if the decision requirement is to form a diagnosis, and if the expected strategy will be feature matching, you will be shown which features will be used, how the features will be detected, what types of inferences will come into play, and so forth.

3. You can request that the decision requirements be part of a more general Cognitive Systems Engineering approach, to include information about workload, memory, and attention.

4. You can expect to use decision requirements to understand how the operators will be reasoning as they make diagnostic and action

decisions. The decision requirements will tell you what to represent, and will not specify how to present the information.

5. You can selectively use Cognitive Task Analysis to identify and represent the decision requirements.

6. You can introduce decision requirements early in the process, during early conceptual design, using scenarios to block out the way operators will be performing their tasks.

7. You can apply decision requirements during preparation of specifications, using analogue cases to suggest concepts for presenting information. Eventually, you may have available pattern books, showing different ways to portray altitude, or changes in the rate of change in a variable.

8. You can apply decision requirements during T&E. Rather than just making sure that the software works as advertised, the T&E phase can be used to conduct a Pre-Mortem, and figure out under what conditions the system will fail to support decision making. The T&E scenarios can be built to reflect these contingencies, to assess whether the operators are able to work their way out of the difficulty.

9. You can apply decision requirements during redesign, to perform Cognitive Task Analysis using actual incidents of system use/misuse.

10. You can use Cognitive Task Analysis, particularly critical incident interviews and controlled observation, to see and experience what the operator is going through. This goes beyond assertions to take the user's needs into account, by giving you a means of understanding the user's choices.

11. You can assemble pattern books showing how different decision requirements have been handled. For instance, decisions about an adversary's intent, or the amount of time remaining versus the lag times in reacting, may occur in a variety of domains. The interface concepts used to handle these types of decisions can be instructive. By seeing what has worked, ideas can be generated. By seeing what has not worked, pitfalls can be avoided. These would be different pattern books from the ones mentioned above in #7 (which would show ways to portray different cues).

12. You can assess the impact of an advanced system by treating it as a team member, to find out where it would be a valuable addition to the team, and where it would be an unacceptable team member.

13. You can consider team decision requirements, looking at the way the proposed work station will affect the way the team members will work with each other.

14. You can assign one member of a design team the role of an Operator's Advocate to track the impact of various tradeoffs on the operator's decision requirements.

The NDM approach is still new. We hope it will trigger much more research and development, with each feeding the other. As we learn more about the nature of decision making in naturalistic settings, we can strengthen the decision requirement process. As we conduct more applications of NDM, we can ask better research questions. As a result of this positive feedback cycle, we should be able to design systems that better support the challenging cognitive tasks.

ENDNOTES

This Endnote Section contains references to related work. We decided not to cite every reference in the text because we didn't want to slow the reader down. On the one hand, we wanted to provide an inclusive set of references, but on the other we didn't want this report to read like a textbook. The sources that seemed less relevant were cited as endnotes to make the information available. The full citation for each source is presented in the reference section.

1. Klein, G.A., & Thordsen, M.L. (1991).

2. Adapted from Orasanu, J., & Connolly, T. (1993).

3. Lussier, J.W. (1991).

4. See Baron, J. (1988); von Winterfeldt, D., & Edwards, W. (1986).

5. Keeney, R.L., & Raiffa, H. (1976).

6. Hammond, K.R., Hamm, R.M., Grassia, J., & Pearson, T. (1987)

7. Raiffa, H. (1968); Edwards, W. (1954, 1962, 1968); Zsambok, Beach, & Klein (1992).

8. See Hammond, K.R. (1993); Woods, D.D. (1993) for some suggestions.

9. Klein, G.A., Orasanu, J., Calderwood, R., & Zsambok, C.E. (1993).

10. Svenson, O. (1979).

11.   The representation was suggested by Len Adelman (personal communication, 1992).

12.   See Klein, G., Wolf, S., Militello, L., & Zsambok, C. (in press).

13.   Janis, I.L., & Mann, L. (1977).

14.   Ibid.

15.   e.g., Isenberg, D.J. (1984, November/December); Klein, G.A. (19891); Soelberg, P.O. (1967).

16.   Klein, G. (1992); Klein, G.A. (1989); Klein, G.A., Calderwood, R., & Clinton-Cirocco, A. (1986).

17.   Brezovic, C.P., Klein, G.A., & Thordsen, M. (1987).

18.   Kaempf, G.L., Wolf, S., Thordsen, M.L., & Klein, G. (1992).

19.   Crandall, B., & Calderwood, R. (1989).

20.   Mosier, K.L. (July, 1990).

21.   Klein, G.A., & Brezovic, C.P. (1986).

22.   Op. cit., Klein et al., 1992.

23.   For more details about all of these models, see Klein, G.A., Orasanu, J., Calderwood, R., & Zsambok, C.E. (1993).

24.   See articles by Crane, P.M. (1992); Endsley, M.R. (1990); Fracker, M.L. (1988).

25. Reising and Moss (1985) have shown the value of building multi-functional interfaces that can be modified depending on the goals being pursued. Depending on the phase of the mission (e.g., ingress, target acquisition, egress), the display can be reconfigured to highlight different cues, and a small set of multi-function buttons would switch accordingly.

26. In the work of Rasmussen, (1990), this corresponds to Rule-Based Behavior. In the field of problem solving, this analogue is the production rule, as in the work of Anderson, Lepper, & Ross, L. (1980).

27. Gentner, D. (1982); Gick, M.L., & Holyoak, K.J. (1980).

28. See Riesbeck, C.K., & Schank, R.C. (1989); Reed, F.D., & Klinger, D.W. (1991).

29. Einhorn, H.J., & Hogarth, R.M. (1981).

30. Tversky, A. (1972).

31. Klein, G. (1992).

32. Edwards (1954).

33. von Winterfeldt, D., & Edwards, W. (1986).

34. Reason, J. (1990); Rasmussen, J.J., Duncan, P.C., & Leplat, J. (1987).

35. See Richard Nisbett and Paul Slovic testimonies before the U.S. House of Representatives Committee on Armed Services on the subject of USS Vincennes downing of Iran Air Flight 655, dated October 6, 1988.

36. Ibid.

37.   Klein, G. (1992).

38.   Klein, G.A., & Thordsen, M.L. (1991).

39.   Klein, G.A., & Crandall, B.W. (in press).

40.   Perrow, C. (1984).

41.   Kahneman, D., Slovic, P., & Tversky, A (Eds.) (1982).

42.   For some other applications see Kleinman and Serfaty
       (1989) whose normative-descriptive framework is built
       around the impact of decision biases for command-and-
       control decisions. Kleinman's framework is a normative
       descriptive framework, to fill in that gap; Silverman
       (1990) speculates on the use of expert systems to catch and
       prevent biases. Wickens and Flach (1987) have
       incorporated biases into an information processing model
       of decision making.

43.   Lopes, L.L. (April 1988).

44.   Gigerenzer, G., & Murray, D.J. (1987).

45.   Cohen, M.S. (1993); Cohen, M.S. (1993).

46.   See Driskell, J.E., & Salas, E. (1991).

47.   For detailed accounts, see reviews by Edland, A. (1989);
       Driskell, J.E. Hughes, S.C., Guy, W., Willis, R.C.,
       Cannon-Bowers, J., & Salas, E. (1991); Mross, E.F., &
       Hammond, K.R. (1990); Hammond, K.R. (1990);
       Hammond, K.R., & Doyle, J.K. (1991); Klein, G., &
       Zsambok, C.E. (in press).

48.   Easterbrook (1959); Edland (1989); Hammond, K.R.
       (1990).

49. de Keyser, V., & Woods, D.D. (in press); Kienan, G. (1987); Staw, R.M., Sandelands, L.E., & Dutton, J.E. (1981).

50. Calderwood, R., Klein, G.A., & Crandall, B.W. (1988); Edland, A. (1989; Hockey, G.R.J. (1970).

51. Op cit., U.S. House of Representative Committee (1988).

52. Ibid.

53. Dotteway, K.A. (1992).

54. Op. cit., Klein & Zsambok (in press).

55. Op. cit., de Keyser & Woods (in press).

56. Klein, G.A., & Brezovic, C.P. (1987).

57. Polanyi, M. (1966).

58. Grant & Mayes (1991); NATO AC/243 (Panel 8/RSG.19) (1991); Redding, R.E. (1990); Woods, D.D. (1993).

59. Gowin, D.B., & Novak, J.D. (1984); McFarren, M.R. (1987); McNeese, M.D., Zapf, B.S., Peio, K.J., Snyder, D.E., Duncan, J.C., & McFarren, M.R. (1990).

60. Cohen, M.S. (1990).

61. Op. cit., Klein (1989); Klein, G.A., Calderwood, R., & MacGregor, D. (1989).

62. Woods (1993).

63. Zachary, W.W., Zaklad, A.L., Hicinbotham, J.H., Ryder, J.M., Purcell, J.A., & Wherry, R.J. (1991).

64.   Smith, P.J., Giffin, W., Rockwell, T., & Thomas, M. (1988).

65.   Brunswik, E. (1966); Hammond, K.R. (1966).

66.   Aldenderfer, R.K. & Blashfield, M.S. (1984).

67.   Fransella, F. & Bannister, D. (1977).

68.   Klein, G.A. (1990).

69.   Andriole, S.J. & Adelman, L. (1989); Rasmussen, J., Pejtersen, A.M., & Goodstein, L. (1991); Woods, D.D. (1993).

70.   Wack, Pierre (September/October, 1985).

71.   Andriole, S.J. (1989); Andriole, S.J. (1990).

72.   Adelman, L. (1992).

73.   See Kaempf, G.L., Wolf, S., Thordsen, M.L., & Klein, G. (1992); Miller, T.E., Wolf, S.P., Thordsen, M.L., & Klein, G. (1992).

74.   Klein, G.A., & Thordsen, M. (1988).

75.   Kaempf, G.L., Klein, G.A., & Thordsen, M.L. (1991).

76.   Eldridge, D., Dodd, R.S., & Mangold, S.J. (1991); Op. cit., Kaempf et al. (1991); Wiener, E.L., Chidester, T.R., Kanki, B.G., Palmer, E.A., Curry, R.E., & Gregorich, S.E. (1991); Woods, D.D., & Sarter, N.B. (1991).

REFERENCES

Adelman, L. (1992). *Evaluating decision support and expert systems*. New York: Wiley.

Aldenderfer, R. K., & Blashfield, M. S. (1984). Cluster analysis. *Sage University paper series on quantitative applications in the social sciences*. Beverly Hills, CA: Sage Publications.

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology, 39*, 1037-1049.

Andriole, S. J. (1989). *Handbook of decision support systems*. Blue Ridge Summit, PA: TAB Books.

Andriole, S. J. (1990). *Information system design principles for the 90s: Getting it right*! Fairfax, VA: AFCEA International Press.

Andriole, S. J., & Adelman, L. (1989). *Cognitive engineering of advanced information technology for Air Force systems design & development* (Final Technical Report AF-AAMRL-HED-SBIR-Phase I). Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.

Baron, J. (1988). *Thinking and deciding*. New York: Cambridge University Press.

Barry, J., & Charles, R. (1992, July). Sea of lies. *Newsweek*, 29-39.

Beach, L. R. (1990). *Image theory: Decision making in personal and organizational contexts*. West Sussex, England: Wiley.

Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *Academy of Management Review*, *3*, 439-449.

Brezovic, C. P., Klein, G. A., & Thordsen, M. (1987). *Decision making in armored platoon command* (AD-A231775). Alexandria, VA: Defense Technical Information Center.

Brunswik, E. (1966). Reasoning as a universal behavior model and a functional differentiation between "perception" and "thinking." In K. Hammond (Ed.), *The psychology of Egon Brunswik* (pp.487-494). New York: Holt, Rinehart, & Winston.

Calderwood, R., Klein, G. A., & Crandall, B. W. (1988). Time pressure, skill, and move quality in chess. *American Journal of Psychology*, *101*, 481-493.

Cannon-Bowers, J. A., Salas, E., & Converse, S. A. (1990). Cognitive psychology and team training: Training shared mental models of complex systems. *Human Factors Society Bulletin*, *33*(12).

Christensen-Szalanski, J. J. J. (1986). Improving the practical utility of judgment research. In B. Brehmer, H. Jungermann, P. Lourens, & G. Sevón (Eds.), *New directions for research in judgment and decision making*. New York: North-Holland.

Cohen, M. S. (1989, August). Decision making "biases" and support for assumption-based higher-order reasoning. *Proceedings of the Fifth Workshop on Uncertainty in Artificial Intelligence*, 71-80. Windsor, Ontario.

Cohen, M. S. (1990). *Conflict resolution as a knowledge elicitation technique* (Working Paper). Reston, VA: Decision Science Consortium.

Cohen, M. S. (1991). Can decision analysis define the goals of training? *Proceedings of the 35th Annual Meeting of the Human Factors Society*, *2*, 1353-1357. Santa Monica, CA: Human Factors Society.

Cohen, M. S. (1993). The naturalistic basis of decision biases. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods*. Norwood, NJ: Ablex.

Cohen, M. S. (1993). Three paradigms for viewing decision biases. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods*. Norwood, NJ: Ablex.

Connolly, T., & Wagner, W. G. (1988). Decision Cycles. In R. L. Cardy, S. M. Puffer, & M. M. Newman (Eds.), *Advances in information processing in organizations*, *3*, 183-205. Greenwich, CT: JAI Press.

Crandall, B., & Calderwood, R. (1989). *Clinical assessment skills of experienced neonatal intensive care nurses* (Final Report for the National Institutes for Health, National Center for Nursing Research, Contract #1 R43 NR01911 01). Yellow Springs, OH: Klein Associates.

Crane, P. M. (1992). *Models of expertise applied to tactical pilot training*. Williams Air Force Base, AZ: U.S. Air Force Armstrong Laboratory.

Crumley, L. M. (1989). *Review of research and methodologies relevant to Army command and control performance measurement* (TR 825). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Crumley, L. M., & Sherman, M. B. (1990, September). *Review of command and control models and theory* (TR 9-15). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

de Groot, A. D. (1946/1965). *Thought and choice in chess* (1st ed.). New York: Mouton.

de Keyser, V., & Woods, D. D. (1993). Fixation errors: Failures to revise situation assessment in dynamic and risky systems. In A. G. Colombo & A. Saiz de Bustamente (Eds.), *Advanced systems in reliability modeling*. Norwell, MA: Kluwer Academic.

Doherty, M. E. (1993). A laboratory scientist's view of naturalistic decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods*. Norwood, NJ: Ablex.

Dotteway, K. A. (1992). *Systematic analysis of complex dynamic systems: The case of the USS Vincennes*. Unpublished master's thesis, Naval Graduate School, Monterey, CA.

Driskell, J. E., Hughes, S. C., Guy, W., Willis, R. C., Cannon-Bowers, J., & Salas, E. (1991). *Stress, stressors, and decision-making* (Technical Report for the U.S. Naval Training Systems Center). Orlando, FL: U.S. Naval Training Systems Center.

Driskell, J. E., & Salas, E. (1991). Overcoming the effects of stress on military performance: Human factors design, training, and selection strategies. In R. Gal & A. D. Mangelsdorff (Eds.), *Handbook of military psychology*. New York: Wiley.

Duffy, L. (1990). Team decision making and group decision support systems. *Proceedings of the 1990 Symposium on Command-and-Control Research*. McLean, VA: Science Applications International Corporation.

Duffy, L. (1993). Team decision making biases: An information processing perspective. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods*. Norwood, NJ: Ablex.

Duncker, K. (1935/1945). On problem solving. *Psychological Monographs*, *58*(5, Whole No. 270).

Dyer, J. L. (1984). Team research and team training: A state-of-the-art review. In F. Muckler (Ed.), *Human Factors Review*. Santa Monica, CA: Human Factors Society.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, *56*,183-201.

Eddy, D. R. (1989). *Selected team performance measures in a C⁴ environment: An annotated bibliography* (Final Report USAF-SAM-TR-87-25). Dayton, OH: Universal Energy Systems.

Edland, A. (1989). On cognitive processes under time stress: A selective review of the literature on time stress and related stress. *Reports from the Department of Psychology*. Stockholm: University of Stockholm.

Edwards, W. (1954). The theory of decision making. *Psychological Review*, *51*, 380-417.

Edwards, W. (1962). Dynamic decision theory and probabilistic information processing. *Human Factors*, *4*, 59-74.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment*. New York: Wiley.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology*, *32*, 53-88.

Eldridge, D., Dodd, R. S., & Mangold, S. J. (1991). *A review and discussion of flight management system incidents reported to the aviation safety reporting system.* Columbus, OH: Battelle Memorial Institute.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning.* Cambridge, MA: Harvard University Press.

Endsley, M. R. (1990). Predictive utility of an objective measure of situation awareness. *Proceedings of the 34th Annual Meeting of the Human Factors Society.* Santa Monica, CA: Human Factors Society.

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. A. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*, (pp. 422-444). Cambridge, England: Cambridge University Press.

Fischhoff, B. (1991). *Giving advice: Decision theory perspectives on sexual assault.* Invited address for the 99th Annual Convention of the American Psychological Association, San Franciso, CA.

Fracker, M. L. (1988). A theory of situation assessment: Implications for measuring situation awareness. *Proceedings of the 32nd Annual Meeting of the Human Factors Society.* Santa Monica, CA: Human Factors Society.

Fransella, F., & Bannister, D. (1977). *A manual for repertory grid technique.* New York: Academic Press.

Fraser, J. M., Smith, P. J., & Smith, J. W. (1990). *A catalog of errors.* Columbus, OH: Ohio State University, Department of Industrial & Systems Engineering.

Gentner, D. (1982). Are scientific analogies metaphors? In D. S. Miall (Ed.), *Metaphor: Problems and perspectives.* Brighton, England: Harvester Press.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*, 306-355.

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3).

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.

Gowin, D. B., & Novak, J. D. (1984). *Learning to learn*. New York: Cambridge University Press.

Grant, S., & Mayes, T. (1991). Cognitive task analysis? In G. R. S. Weir & J. L. Alty (Eds.), *Human computer interaction and complex systems*. London: Academic Press.

Hammond, K. R. (1966). Probabilistic functionalism: Egon Brunswik's integration of the history, theory, and methodology of psychology. In K. R. Hammond (Ed.), *The psychology of Egon Brunswik*. New York: Holt, Rinehart, & Winston.

Hammond, K. R. (1990). *The effects of stress on judgement and decision making: An overview and arguments for a new approach*. Boulder, CO: University of Colorado, Center for Research on Judgment and Policy.

Hammond, K. R. (1993). Naturalistic decision making from a Brunswikian viewpoint: Its past, present, future. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods*. Norwood, NJ: Albex.

Hammond, K. R., & Doyle, J. K. (1991). *Effects of stress on judgment and decision making: Part II* (Prepared for the U.S. Army Research Office). Research Triangle Park, NC: U.S. Army Research Office.

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Transactions on Systems, Man, and Cybernetics*, *17*(5), 753-770.

Hockey, G. R. J. (1970). Effect of loud noise on attentional selectivity. *Quarterly Journal of Experimental Psychology*, *22*, 28-36.

Isenberg, D. J. (1984, November/December). How senior managers think. *Harvard Business Review*, 80-90.

Janis, I. L., & Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment*. NY: Free Press.

Kaempf, G. L., Klein, G. A., & Thordsen, M. L. (1991). *Applying recognition-primed decision making to man-machine interface design* (Final report for NASA/Ames Research Center, Moffett Field, CA, Contract NAS2-13359). Fairborn, OH: Klein Associates.

Kaempf, G. L., Wolf, S., Thordsen, M. L., & Klein, G. (1992). *Decision making in the AEGIS combat information center* (Prepared for the U.S. Naval Command, Control and Ocean Surveillance Center, San Diego, CA, Contract N66001-90-C-6023). Fairborn, OH: Klein Associates.

Kahan, J. P., Worley, D. R., & Stasz, C. (1989). *Understanding commander's information needs* (R-3761-A). Santa Monica, CA: Rand Corporation.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, England: Cambridge University Press.

Keeney, R. L., & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs*. New York: Wiley.

Kienan, G. (1987). Decision making under stress: Scanning of alternatives under controllable and uncontrollable threats. *Journal of Personality and Social Psychology*, *52*, 639-644.

Klayman, J., & Ha, Y-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*(2), 211-228.

Klein, G. A. (1989). Recognition-primed decisions. In W.B. Rouse (Ed.), *Advances in Man-Machine System Research*, (Vol.5, pp. 47-92). Greenwich, CT: JAI Press.

Klein, G. A. (1990). Knowledge engineering: Beyond expert systems. *Information and Decision Technologies*, *16*(1), 27-41.

Klein, G. (1992). *Decisionmaking in complex military environments*. (Prepared for the U.S. Naval Command, Control and Ocean Surveillance Center, San Diego, CA, Contract N66001-90-C-6023). Fairborn, OH: Klein Associates.

Klein, G. (1992). *Errors in naturalistic decision making*. (Prepared for the U.S. Army Research Institute, Alexandria, VA and the U.S. Naval Command, Control and Ocean Surveillance Center, San Diego, CA, Contracts MDA903-89-C-0032 and N66001-90-C-6023). Fairborn, OH: Klein Associates.

Klein, G. A., & Brezovic, C. P. (1986). Design engineers and the design process: Decision strategies and human factors literature. *Proceedings of the 30th Annual Meeting of the Human Factors Society*, *2*, 771-775. Santa Monica, CA: Human Factors Society.

Klein, G. A., & Brezovic, C. P. (1987). Analytical versus recognitional approaches to design decision making. In W. B. Rouse & K. R. Boff (Eds.), *System design: Behavioral perspectives on designers, tools, and organizations* (pp. 175-186). New York: Elsevier Science Publishing Company.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground, *Proceedings of the 30th Annual Meeting of the Human Factors Society*, *1*, 576-580. Santa Monica, CA: Human Factors Society.

Klein, G. A., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics*, *19*(3), 462-472.

Klein, G. A., & Crandall, B. W. (in press). The role of mental simulation in naturalistic decision making. In J. Flach, P. Hancock, J. Caird, & K. Vicente (Eds.), *The ecology of human-machine systems*. Hillsdale, NJ: Erlbaum.

Klein, G. A., & Crandall, B. W. (1992). *Recognition-primed decision strategies* (Final Technical Report for the U.S. Army Research Institute, Office of Basic Research, Alexandria, VA, Contract MDA903-89-C-0032). Fairborn, OH: Klein Associates.

Klein, G. A., Orasanu, J., Calderwood, R., & Zsambok, C. E. (1993). *Decision making in action: Models and methods*. Norwood, NJ: Ablex.

Klein, G. A., & Thordsen, M. (1988). Use of progressive deepening in battle management. *Proceedings of the 11th Biennial DoD Psychology Conference*. Colorado Springs, CO: U.S. Air Force Academy, Department of Behavioral Sciences and Leadership.

Klein, G. A., & Thordsen, M. L. (1991). Representing cockpit crew decision making. *Proceedings of the Sixth International Symposium on Aviation Psychology*. Columbus, OH: Ohio State University, Department of Aviation.

Klein, G. A., & Weitzenfeld, J. (1978). Improvement of skills for solving ill-defined problems. *Educational Psychologist*, *13*, 31-41. (Also released as Wright-Patterson Air Force Base Technical Report AFHRL-TR-78-31.)

Klein, G., Wolf, S., Militello, L., & Zsambok, C. (in press). Characteristics of skilled option generation in chess. Manuscript submitted for publication.

Klein, G., & Zsambok, C. E. (in press). The effect of acute stressors on decision making. In J. E. Driskell & E. Salas (Eds.), *Stress and human performance*. Hillsdale, NJ: Erlbaum.

Kleinman, D. L. & Serfaty, D. (1989, April). Team performance assessment in distributed decision making. In R. Gilson, J. P. Kincaid, & B. Goldiez (Eds.), *Proceedings of the Interactive Networked Simulation for Training Conference*. Orlando, FL.

Klinger, D.W., Andriole, S.J., Militello, L.G., Adelman, L., Klein, G., & Gomes, M.E. (1993). *Designing for performance: A cognitive systems engineering approach to modifying an AWACS human-computer interface* (Final Report for the U.S. Air Force Armstrong Laboratory, Wright-Patterson Air Force Base, OH, Contract F33615-90-C-0533). Fairborn, OH: Klein Associates.

Kyne, M., Militello, L., & Klein, G. (1992). *The effects of mental simulation on confidence*. Fairborn, OH: Klein Associates.

Lipshitz, R. (1989). *Making sense of decision making: Towards an integrative approach to understanding and improving decisions*. Haifa, Israel: University of Haifa.

Lopes, L. L. (April, 1988). *The rhetoric of irrationality*. Madison, WI: University of Wisconsin.

Lussier, J. W. (1991). Two unnatural methods of group problem solving. *Proceedings of the Military Testing Association*. San Diego, CA: U.S. Naval Personnel Research and Development Center.

McFarren, M. R. (1987). *Using concept mapping to define problems and identify key kernels during the development of a decision support system.* Unpublished master's thesis, Air Force Institute of Technology, School of Engineering, Wright-Patterson Air Force Base, OH.

McNeese, M. D., Zapf, B. S., Peio, K. J., Snyder, D. E., Duncan, J. C., & McFarren, M. R. (1990). *An advanced knowledge and design acquisition methodology: Application for the pilot's associate.* (Final Report AAMRL-TR-90-060). Wright-Patterson Air Force Base, OH: U.S. Air Force Armstrong Aerospace Medical Research Laboratory.

Means, B., & Gott, S. P. (1988). Cognitive task analysis as a basis for tutor development: Articulating abstract knowledge representations. In J. Psotka, L. D. Massey, & S. A. Mutter (Eds.), *Intelligent tutoring systems: Lessons learned*, (pp. 35-57). Hillsdale, NJ: Erlbaum.

Miller, T., Wolf, S., Thordsen, M. L., & Klein, G. (1992). *A decision-centered approach to storyboarding anti-air warfare interfaces* (Prepared for the U.S. Naval Command, Control & Ocean Surveillance Center, San Diego, CA., Contract N66001-90-C-6023). Fairborn, OH: Klein Associates.

Montgomery, H. (1983). Decision rules and the search for a dominance structure: Towards a process model of decision making. In P. Humphreys, O. Svenson, & A. Vari (Eds.), *Advances in psychology*, (pp. 343-369). New York: North-Holland.

Morgan, G. (1986). *Images of organization.* London: Sage Publications.

Mosier, K. L. (July, 1990). *Decision making on the air transport flight deck: Process and product.* Unpublished doctoral dissertation, University of California, Berkeley, CA.

Mross, E. F., & Hammond, K. R. (1990). The effects of stress on judgment and decision making: An annotated bibliography (Report No. 295). Boulder, CO: University of Colorado, Center for Research on Judgment and Policy.

NATO Defense Research Group AC/243 (Panel 8/RSG.19) (1991). *Toward a framework for cognitive analysis, design, and evaluation (COADE)*. Washington, DC. Author.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NY: Prentice-Hall.

Noble, D. (1989, January). *Application of a theory of cognition to situation assessment* (Technical report for the U.S. Office of Naval Research, Contract N00014-84-0484). Vienna, VA: Engineering Research Associates.

Noble, D., Boehm-Davis, D., & Grosz, C. G. (1986). *A schema-based model of information processing for situation assessment*. Vienna, VA: Engineering Research Associates. (NTIS No. ADA163150)

Olmstead, J. A. (1990). *Battle staff integration*. Ft. Leavenworth, KS: U.S. Army Command and General Staff College.

Orasanu, J.M. (1990). *Shared mental models and crew decision making*. Princeton, NJ: Princeton University, Cognitive Science Laboratory.

Orasanu, J., & Connolly, T. (1993). The reinvention of decision making. In G. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods*. Norwood, NJ: Albex.

Orasanu, J., & Salas, E. (1993). Team decision making in complex environments. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods*. Norwood, NJ: Ablex.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance, 16*, 366-387.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 534-552.

Pennington, N., & Hastie, R. (1981). Juror decision making models: The generalization gap. *Psychological Bulletin, 89*, 246-287.

Pennington, N., & Hastie, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 14*(3), 521-533.

Perrow, C. (1984). *Normal accidents: Living with high-risk technologies.* New York: Basic Books.

Pew, R. W., Miller, D. C., & Feehrer, C. E. (1982). *Evaluation of Proposed Control Room Improvements Through Analysis of Critical Operator Decisions* (NP-1982). Palo Alto, CA: Electric Power Research Institute.

Polanyi, M. (1966). *The tacit dimension.* Garden City, NY: Doubleday.

Politser, P. E. (1989). Cognitive guidelines for simplifying medical information: Data framing and perception. *Journal of Behavioral Decision Making, 2*, 149-165.

Raiffa, H. (1968). *Decision analysis: Introductory lectures on choices under uncertainty.* Reading, MA: Addison Wesley.

Rasmussen, J. (1985). The role of hierarchical knowledge representation in decision making and system management. *IEEE Transactions on Systems, Man and Cybernetics, 15*(2), 234-243.

Rasmussen, J. (1990). The role of error in organizing behaviour. *Ergonomics, 33*(10/11), 1185-1199.

Rasmussen, J. J., Duncan, P. C., & Leplat, J. (1987). *New technology and human error*. New York: Wiley.

Rasmussen, J. J., Pejtersen, A. M., & Goodstein, L. (1991). *Cognitive engineering: Concepts and applications* (2nd ed.). New York: Wiley.

Reason, J. (1990). *Human error*. Cambridge, England: Cambridge University Press.

Redding, R. E. (1990). Taking cognitive task analysis into the field bridging the gap from research to application. *Proceedings of the 34th Annual Meeting of the Human Factors Society, 2*, 1304-1308. Santa Monica, CA: Human Factors Society.

Reed, F. D., & Klinger, D. W. (1991). *A case-based reasoning system for performing manufacturing assessments: Phase II* (Final Technical Report for ASD/PMRRC Wright-Patterson Air Force Base, OH, Contract F33615-89-C-5702). Fairborn, OH: Klein Associates.

Reising, J. M., & Moss, R. W. (1985, October). Cockpit automation: Design and implementation issues. *Proceedings of the Society of Automotive Engineers Aerospace Technology Conference and Exposition*. Warrendale, PA: Society of Automotive Engineers.

Riesbeck, C. K., & Schank, R. C. (1989). *Inside case-based reasoning*. Hillsdale, NJ: Erlbaum.

Rogers, W., & Rogers, S. (1992). *Storm center: The USS Vincennes and Iran flight 655*. Annapolis, MD: Naval Institute Press.

Russo, J. E., & Shoemaker, P. J. H. (1989). *Decision traps*. Garden City, NY: Doubleday.

Sage, A. P. (1981). Behavioral and organizational considerations in the design of information systems and processes for planning and decision support. *IEEE Transactions on Systems, Man, and Cybernetics, 11*, 640-678.

Segal, L. (1989). *Effects of aircraft cockpit design on crew communication.* Internal paper prepared for the University of Illinois, Urbana-Champaign.

Shanteau, J. (1989). Cognitive heuristics and biases in behavioral auditing: Review, comments, and observations. *Accounting Organizations and Society, 14*(1-2), 165-177.

Shanteau, J., & Phelps, R. H. (1977). Judgment and swine: Approaches and issues in applied judgment analysis. In M. F. Kaplan & S. Schwartz (Eds.), *Human judgment & decision processes in applied settings.* New York: Academic Press.

Silverman, B. G. (1990). Critiquing human judgment using knowledge—acquisition systems. *AI Magazine*, 60-79.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*, 99-118.

Smith, P. J., Giffin, W., Rockwell, T., & Thomas, M. (1988). Modeling fault diagnosis as the activation and use of a frame system. *Human Factors, 28*(6), 703-716.

Soelberg, P. O. (1967). Unprogrammed decision making. *Industrial Management Review, 8*, 19-29.

Staw, R. M., Sandelands, L. E., & Dutton, J. E. (1981). Threat-rigidity effects in organizational behavior: A multi-level analysis. *Administrative Science Quarterly, 26*, 501-524.

Svenson, O. (1979). Process descriptions in decision making. *Organizational Behavior and Human Performance, 23*, 86-112.

Swezey, R. W., & Salas, E. (1992). *Teams: Their training and performance*. Norwood, NJ: Ablex.

Thordsen, M. L., & Klein, G. A. (1989). Cognitive processes of the team mind. *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*. New York: Institute of Electrical and Electronic Engineers.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review, 79*(4), 281-299.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Tversky, A., & Kahneman, C. (1974). On the psychology of prediction. *Psychological Review, 80*, 237-251.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*, 293-315.

von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. Cambridge, England: Cambridge University Press.

Wack, P. (1985, September/October). Scenarios: Uncharted waters ahead. *Harvard Business Review*, 73-89.

Wegner, D. (1987). Transactive memory: A contemporary analysis of group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185-208). New York: Springer-Verlag.

Wickens, C. D., & Flach, J. (1987). Human information processing. In E. L. Wiener & D. C. Nagel (Eds.), *Human factors in aviation*. New York: Wiley.

Wiener, E. L., Chidester, T. R., Kanki, B. G., Palmer, E. A., Curry, R. E., & Gregorich, S. E. (1991). *The impact of cockpit automation on crew coordination and communication: 1. Overview, LOFT evaluations, error severity, and questionnaire data* (NASA Contract Report 177587). Moffett Field, CA: NASA, Ames Research Center.

Wohl, J. C. (1981). Force management decision requirements for Air Force Tactical Command and Control. *IEEE Transactions on Systems, Man, and Cybernetics*, 618-639.

Woods, D. D. (1993). Process tracing methods for the study of cognition outside of the experimental psychology laboratory. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods*. Norwood, NJ: Ablex.

Woods, D. D. (in press). *Beyond the purely visual and decorative in computer-based information displays: The theory and practice of representation design in a computer medium*. Columbus, OH: Ohio State University.

Woods, D.D. (1990). Modeling and predicting human error. In J. Elkind, S. Card, J. Hochberg, & B. Huey (Eds.), *Human performance models for computer-aided engineering*. New York: Academic Press.

Woods, D. D., & Sarter, N. B. (1991). Pilot interaction with cockpit automation. Presented at the *Second AMES AS/A Program Investigator's Meeting*.

Zachary, W. W., Zaklad, A. L., Hicinbotham, J. H., Ryder, J. M., Purcell, J. A., & Wherry, R. J. (1991). *COGNET representation of tactical decision-making in ship-based anti-air warfare*. (Technical Report 911015.9009). Spring House, PA: CHI Systems.

Zsambok, C. E., Beach, L. R., & Klein, G. (1992). *A literature review of analytical and naturalistic decision making.* (Prepared for the Naval Command, Control and Ocean Surveillance Center, San Diego, CA, Contract N66001-90-C-6023). Fairborn, OH: Klein Associates.

Zsambok, C. E., Klein, G., Kyne, M. M., & Klinger, D. W. (1992). *Advanced team decision making: A developmental model.* (Prepared for the U.S. Army Research Institute for the Behavioral and Social Sciences, Contract MDA903-90-C-0117 (DTIC No. AD-A259512)) Fairborn, OH: Klein Associates Inc.

AUTHOR INDEX

SUBJECT INDEX